The international journal of science / 20 August 2020

nature

COVID QUESTIONS

From virus transmission to antibody response

Catch-22
Why developing new antibiotics can be bad for business

Shrinking structures
3D printing allows miniaturization of silica aerogels

Reptiles revisited
Genome sequence offers evolutionary insight into the tuatara

The international journal of science / 20 August 2020

nature

Progress report on a pandemic

In the first of a series of editorials, we look back at some of the key findings from scientists' race to demystify the new coronavirus.

In the space of eight months, the new coronavirus SARS-CoV-2 and the disease it causes, COVID-19, have dominated the work of thousands of researchers in an unprecedented global effort.

In a series of editorials, we look back at key scientific findings that have revealed important characteristics of the virus and COVID-19, including emerging approaches to treatment and prevention. We begin, this week, with how the virus was identified; how it transmits between people; and the many ways in which it affects the human body.

Cracking the virus code

When an outbreak of a disease similar to severe acute respiratory syndrome (SARS) emerged in Wuhan, China, at the end of 2019, scientists suspected that a new coronavirus had spread to humans. Many of the first cases to be identified were linked to a single live-animal market in the city.

Researchers in China immediately began working to isolate and sequence the virus. When the original SARS virus, now known as SARS-CoV-1, emerged in humans in 2002, it took months to obtain a full sequence of the virus genome. This time, advances in sequencing technologies meant that scientists were able to unpick the virus's RNA code within weeks of the first cases appearing.

On 11 January, Yong-Zhen Zhang at Fudan University in Shanghai and his colleagues deposited the genome sequence of a virus isolated from a 41-year-old who had worked at the animal market into a public database. In doing so, they alerted the world to the existence of a new coronavirus that was related to SARS-CoV-1. Their findings were subsequently published in *Nature*[1].

Although Zhang's team had sequenced the virus from only a single patient, simultaneous work by other groups identified the same virus from other people with pneumonia. Together, these researchers firmly implicated this new coronavirus as the cause of the disease. One of the teams, led by Shi Zhengli at the Wuhan Institute of Virology, also determined that the closest known relative of the new virus was a bat coronavirus[2].

Not just a respiratory virus

Initial reports of the disease, named COVID-19 on 11 February, described a severe respiratory illness similar to that caused by SARS-CoV-1. Chest scans showed patchy shadows – known as 'ground glass opacities' – in the lungs of many patients, according to early studies from hospitals in Wuhan[3]. Moreover, older people, men and those with other diseases were more likely to be admitted to intensive care, whereas children seemed to have milder disease[4].

But it quickly became apparent that SARS-CoV-2 is not just a respiratory virus. It also affects blood vessels, causing thrombosis[5] and strokes[6].

Autopsies have found the virus in organs other than the lungs, including the kidneys, liver, heart and brain, as well as in the blood[7]. We now know that symptoms of COVID-19 can include gastrointestinal, neurological, renal, cardiovascular and other complications[8].

> "It quickly became apparent that SARS-CoV-2 is not just a respiratory virus."

Something in the air

It soon became clear that SARS-CoV-2 could hop from one person to another. This could happen through direct contact or indirect transmission, such as through droplets expelled during a cough, or even a simple exhalation. What wasn't clear – and is still a matter of debate – is how big those droplets need to be, and how far they can travel.

It's an important question. Larger droplets will quickly fall to the ground, but smaller, lighter ones – known as aerosols – can stay suspended in the air. A virus that can hitch a ride on such tiny droplets can travel farther and could raise the risk of infection in poorly ventilated indoor spaces.

The potential of the new coronavirus to travel in this way was the focus of a study, published in April, on SARS-CoV-2 aerodynamics in two hospitals in Wuhan[9]. Researchers found that some areas of the hospitals, particularly some staff areas, had relatively high concentrations of viral RNA in aerosol-sized droplets. The team did not determine whether those droplets were infectious.

Invisible disease

As the virus began to spread around the world, there were suggestions that people without symptoms might be able to transmit it.

In March, data from the cruise ship *Diamond Princess* revealed that 17.9% of those who tested positive for COVID-19 on the ship had no symptoms[10]. More than 3,700 people had been quarantined aboard the vessel in February after a former passenger was found to have COVID-19. In April, a study of 94 people showed that 'viral shedding' – the release of a virus into the environment – seemed to peak before or at the same time as the onset of symptoms[11].

We have come a long way in understanding how the pandemic arose and how it spread around the world – by studying the virus's characteristics and transmission, and how it causes disease. In future instalments of this editorial series, we'll look at the research on how to control it, as well as progress on treatments and vaccines.

1. Wu, F. *et al. Nature* **579**, 265–269 (2020).
2. Zhou, P. *et al. Nature* **579**, 270–273 (2020).
3. Huang, C. *et al. Lancet* **395**, P497–P506 (2020).
4. Lu, X. *et al. N. Engl. J. Med.* **382**, 1663–1665 (2020).
5. Ackermann, M. *et al. N. Engl. J. Med.* **383**, 120–128 (2020).
6. Oxley, T. J. *et al. N. Engl. J. Med.* **382**, e60 (2020).
7. Puelles, V. G. *et al. N. Engl. J. Med.* **383**, 590–592 (2020).
8. Gupta, A. *et al. Nature Med.* **26**, 1017–1032 (2020).
9. Liu, Y. *et al. Nature* **582**, 557–560 (2020).
10. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. *Eurosurveillance* **25**, 2000180 (2020).
11. He, X. *et al. Nature Med.* **26**, 672–675 (2020).

World view

By Nicole Mather

How we accelerated clinical trials in the age of COVID-19

The United Kingdom's RECOVERY trial shows a way to benefit patients faster.

In March, as the tsunami of COVID-19 hit Europe, it became obvious that the virus could overwhelm the United Kingdom's National Health Service (NHS). To address this issue, colleagues and I repurposed infrastructure so that clinical trials could safely get data about more treatments from more patients more quickly.

This allowed the NHS to run the biggest randomized COVID-19 clinical trial in the world – and to identify a treatment, amid the heat of the epidemic, without bypassing regulatory processes. It built on investment in programmes and infrastructure established in 2017 as government strategy, when I was director of the Office for Life Sciences.

We worked nights and weekends to pivot NHS DigiTrials services – which had been set up in 2019 for planning large clinical trials – towards providing more kinds of information, including patient results, and applied the new services to the ambitious RECOVERY trial. This trial, based at the University of Oxford, aims to rapidly test a range of potential treatments for people ill with COVID-19. If any such treatments work, moving faster could save more lives.

On 16 June , RECOVERY announced that dexamethasone, a commonly available steroid, could reduce mortality by one-third in people with severe respiratory complications owing to COVID-19. Remarkably, this study encompassed 12,000 patients and 176 sites over a 3-month period. Looking back, I see ideas that could be broadly applied to accelerate trials around the world.

The RECOVERY trial had five key features that distinguish it from a standard approach. It had a short, flexible protocol – just 20 pages long – that laid out the design and data and regulatory requirements, and allowed trial arms to be halted or added. It received ethical and regulatory approval in just 9 days, compared with the standard 30–60 days. Its recruitment procedures were straightforward, with only a two-page consent form and a one-page bedside form to be completed by clinicians. It accelerated data collection and processing through NHS DigiTrials. And it quickly made results public – the announcement was followed by a preprint on the medRxiv server and journal publication within a month (The RECOVERY Collaborative Group. *N. Engl. J. Med.* http://doi.org/gg5c8p; 2020).

What lessons can be applied to trials in the future? How can we revamp procedures and leverage technology to accelerate findings, and do so without sacrificing transparency, patient involvement and peer review?

First, streamline bureaucracy. We've gone so far towards managing risk that we've created layers of bureaucracy that absorb time and money, and, paradoxically, increase the risk that beneficial treatments are not tested – or worse, that ineffective treatments are used widely in the rush to 'do something'. Clinical-trial protocols, ethical-consent forms and patient-information leaflets can run to thousands of words. Review processes can take months, requiring different data sets and sequential approvals.

> "During COVID-19, the Health Research Authority reduced the average ethical-review cycle from 60 to 10 days."

There is no excuse – we must pare down to the key questions to accelerate the process. Some early lessons came during the West Africa Ebola outbreak. During Ebola, and again during COVID-19, the UK Medicines and Healthcare Products Regulatory Agency (MHRA) prioritized and processed clinical-trial applications within a week. During COVID-19, the Health Research Authority (HRA) reduced the average ethical-review cycle from 60 to 10 days.

In the longer term, any approach to prioritization needs careful consideration and consultation, but coordinating regulatory functions can accelerate the process. For example, the Combined Ways of Working pilot programme, launched in 2018, allows clinical-trial applications to be submitted for concurrent review by the MHRA and HRA.

Second, leverage data systems. The RECOVERY trial benefited from UK investments in NHS health-data systems. That includes the work of our NHS DigiTrials team – a consortium of NHS Digital, my team at IBM, the University of Oxford and Microsoft. These data systems meant that only minimal demographic and consent data had to be collected at a patient's bedside and were then integrated with routine NHS information on treatment, diagnosis, COVID-19 tests, clinical results and survival.

Third, enable trust. Accelerating research during COVID-19 meant less opportunity to engage patients in the design and delivery of trials. As trials restart, we must broaden efforts to involve patients and the public. To engender trust in the use of health data for research, and to explain its potential to transform care, we need to work with institutions in which the public has confidence, such as charities or non-governmental organizations.

Fourth, maintain transparency. RECOVERY aimed to balance rapid sharing and expert review. The full protocol and core documents are available on a public website. Key results were made available through public statements, and fuller details were published as preprints simultaneously with submission to a peer-reviewed journal. Results were shared with major international groups such as the World Health Organization. NHS hospitals were urged to adopt the use of dexamethasone within hours of the public announcement.

All these lessons are broadly applicable to many countries. As we turn our attention back to other major causes of illness and death – such as cancer, and cardiovascular and neurodegenerative diseases – we should apply the lessons from COVID-19 to streamline clinical trials and deliver effective treatments.

Nicole Mather is life-sciences lead at IBM Services in London; executive board member at NHS DigiTrials; and non-executive director at the UK Health Research Authority and Wellcome Genome Research Ltd.
e-mail: nicole.mather@ibm.com

MATTHEW JAMES

News in brief

HUGE RADIO TELESCOPE DAMAGED BY CABLE BREAKAGE

The 305-metre-wide dish of the Arecibo Observatory in Puerto Rico — one of the world's pre-eminent radio telescopes — has been damaged by a cable that broke unexpectedly. The cause of the breakage on 10 August is currently unknown, and astronomical observations at the facility have been suspended indefinitely until the damage can be repaired.

One end of the cable slipped out of its socket in the middle of the night and fell, smashing around 250 of the 40,000 panels that make up the main dish and leaving a 30-metre gash. Engineers are investigating what went wrong. The 8-centimetre-thick cable is one of several installed more than two decades ago, and had been expected to last for another 15–20 years.

Observatory director Francisco Córdova said in a press briefing that it wasn't yet clear whether several natural disasters that have ravaged Puerto Rico — including Hurricane Maria in 2017 and a magnitude-6.4 earthquake in January this year — contributed to the failure. "Our commitment is to get this back up and running as quickly as possible," he said. The Arecibo dish typically observes a wide range of astronomical phenomena, including the cosmic flashes known as fast radio bursts, and asteroids that are potentially hazardous to Earth.

FIRST EVIDENCE THAT ANTIBODIES PROTECT AGAINST SARS-COV-2 REINFECTION

A COVID-19 outbreak on a US fishing boat has provided what scientists say is the first direct evidence that antibodies against the new coronavirus protect people from reinfection.

After a viral infection, the immune system makes compounds called neutralizing antibodies that can attack the virus if it invades again. But previous research had not determined whether such antibodies can shield humans from SARS-CoV-2 reinfection.

Alexander Greninger at the University of Washington School of Medicine in Seattle and his colleagues tested the crew of a fishing vessel for SARS-CoV-2 and for antibodies (A. Addetia *et al.* Preprint at medRxiv http://doi.org/d6qm; 2020). Before the ship's departure, the researchers tested 120 of the 122 crew members and found that all were negative for SARS-CoV-2, but an outbreak hit the ship soon after departure.

Post-voyage testing showed that 104 members of the crew were infected. None of those who were infected and had been tested before embarking had shown neutralizing antibodies against SARS-CoV-2.

However, all three crew members who did have such antibodies before departure escaped infection.

2019 AMONG THE THREE HOTTEST YEARS ON RECORD

An international review of the world's climate has found that 2019 was one of the three hottest years on record.

The mean annual global surface temperature last year was about half a degree above the 1981–2010 average, according to the most recent annual *State of the Climate* report, which was compiled by scientists with the National Oceanic and Atmospheric Administration (NOAA) and released on 12 August.

The global concentration of heat-trapping greenhouse gases in the atmosphere climbed to a record high of almost 410 parts per million in 2019, which in turn led to a record number of extremely warm days. The year also had the second-highest average global sea surface temperature on record, surpassed only by 2016, when there was an El Niño warming event, the report says.

Although last year was among the hottest on record, its exact rank depends on the data set used. According to data from NOAA and NASA, 2019 was the second-hottest year since records began in the nineteenth century. The UK Met Office, which runs independent climate measurements, lists last year as the third-hottest on record, behind 2016 and 2015.

The report notes that, regardless of which historical data set is used, the six warmest years on record have all been in the past six years.

Meanwhile, it is possible that 2020 has set a new heat record already. A temperature of 54.4 °C was recorded in Death Valley in eastern California (pictured) on 16 August. If this measurement is confirmed, it will be the highest air temperature observed on Earth in more than a century.

L TO R: ARECIBO OBSERVATORY/UNIV. CENTRAL FLORIDA; MATT STONE/MEDIANEWS GROUP/*BOSTON HERALD* VIA GETTY; JASON OGULNIK FOR *NATURE*

News in focus

MEL EVANS/AP/SHUTTERSTOCK

Princeton University will rename a programme to remove association with Woodrow Wilson, who discouraged enrolment of Black students.

UNIVERSITIES SCRUB NAMES OF RACIST LEADERS — STUDENTS SAY IT'S A FIRST STEP

Activists are glad to see progress, but now call for deeper cultural change in academia.

By Giuliana Viglione & Nidhi Subbaraman

Nearly five years ago, the Black Justice League student group at Princeton University in New Jersey organized a sit-in at the office of the institution's president to demand that Woodrow Wilson's name be removed from its vaunted public-policy programme.

When he was president of Princeton from 1902 to 1910, Wilson discouraged the enrolment of Black students, and as president of the United States from 1913 to 1921, he supported segregating white and Black employees in the federal government. Although the 2015 sit-in didn't convince Princeton's trustees to wipe Wilson's name, this year's wave of demonstrations against racism prompted action. The protests, sparked when George Floyd was killed by police in Minneapolis, Minnesota, in May, are part of the Black Lives Matter movement, which calls for an end to police violence and systemic racism against Black people. In June, Princeton announced that it would rename the programme, as well as a residential college.

The university is not alone in rethinking its legacy. In June, the University of Southern California (USC) in Los Angeles removed a former president's name from a central campus building because he supported eugenics. In the same month, the University of Mons in Belgium removed a bust of Leopold II, the Belgian king who at the turn of the twentieth century led a brutal and bloody colonial campaign in what is now the Democratic Republic of the Congo. And in July, Cold Spring Harbor Laboratory in New York removed DNA scientist James Watson's name from its biological-sciences graduate programme, citing his past racist comments.

The Black Lives Matter movement has spurred institutions worldwide to announce that they will change or review the names of campus buildings, programmes and memorials dedicated to scientists and other figures who had discriminatory beliefs. Many of these announcements followed years-long campaigns by students and faculty members who risked their careers to remake their institutions from within. "We got to a tipping point," says Susan Reverby, a historian of medicine who studies equality and ethics in public health at Wellesley College in Massachusetts. "But we wouldn't have gotten to the tipping point if people hadn't done all the work they've been doing for generations to try to fight this."

Still, those who fought for the changes say that renaming buildings is only the first step towards improving diversity and inclusion in academia; they are advocating sustained efforts to transform university culture.

Delayed action

Like Princeton, many of the institutions that have recently renamed buildings and memorials had earlier opportunities to do so and didn't take them.

"It's not that Princeton changed its mind, it's that public opinion changed around them," says Abyssinia Lissanu, a graduate student in public policy who is part of the Princeton Policy School Demands group, one of several that have been pressuring the administration to make the university more inclusive.

In February, University College London (UCL) committed to dropping the names of Francis Galton and Karl Pearson, celebrated statisticians who supported eugenics, from buildings and lecture halls on campus. "Then there was a long pause and nothing happened," says Michael Sulu, a UCL biochemical engineer who campaigned for the removal of the names.

According to a university spokesperson, the COVID-19 pandemic delayed action. After George Floyd died and worldwide protests erupted, UCL announced on 19 June that three spaces would have Galton's and Pearson's names removed immediately. They now bear generic names such as Lecture Theatre 115. Sulu credits student groups at the university with keeping up the pressure to ensure change.

Similarly, USC convened a task force last year to re-evaluate its campus buildings and memorials. At the top of the list was the Von KleinSmid Center, one of the university's most prominent buildings. The centre, which houses the department of international relations, was named after past USC president Rufus Von KleinSmid, who was a member of the now-defunct Human Betterment Foundation, a eugenics organization in southern California that advocated the forced sterilization of people with disabilities. Students had been campaigning for the building to be renamed for years. On 10 June, the university abruptly removed letters spelling out Von KleinSmid's name and a bust of the scientist from the building.

> **"The treatment of the people in the institutions matters just as much as the name that's on them."**

The recent protests haven't sparked swift change everywhere. In February, a student organization at Stanford University in California delivered a formal request that the institution's leaders rename Jordan Hall, which houses its psychology department. The building is named after Stanford's founding president, David Starr Jordan, a marine biologist and famous eugenicist. The psychology faculty delivered its own request with unanimous support for the move the following month. Stanford's naming-review committee says it won't deliver its recommendations until the beginning of the autumn term, although it announced last month that the evaluation was being expedited.

Stanford University's psychology department commemorates David Starr Jordan, a eugenicist.

ARCHITECTOUR/ALAMY

At Stanford, faculty members were instrumental in driving action. Irene Newton, a microbiologist at Indiana University Bloomington (IUB) who co-authored a June petition to rename an IUB building also named after Jordan, says that this is the first time faculty members at her institution have coalesced around the issue, despite previous actions by students. As a faculty member, "you need to look at the power you have and try and make the change you can", she says.

Chris Jackson, a geoscientist at Imperial College London, agrees that faculty members should put their weight behind such efforts. "You have to kind of stand for something. For me, at least, as a professor at a fancy university, what are you going to use your platform for and your position for?"

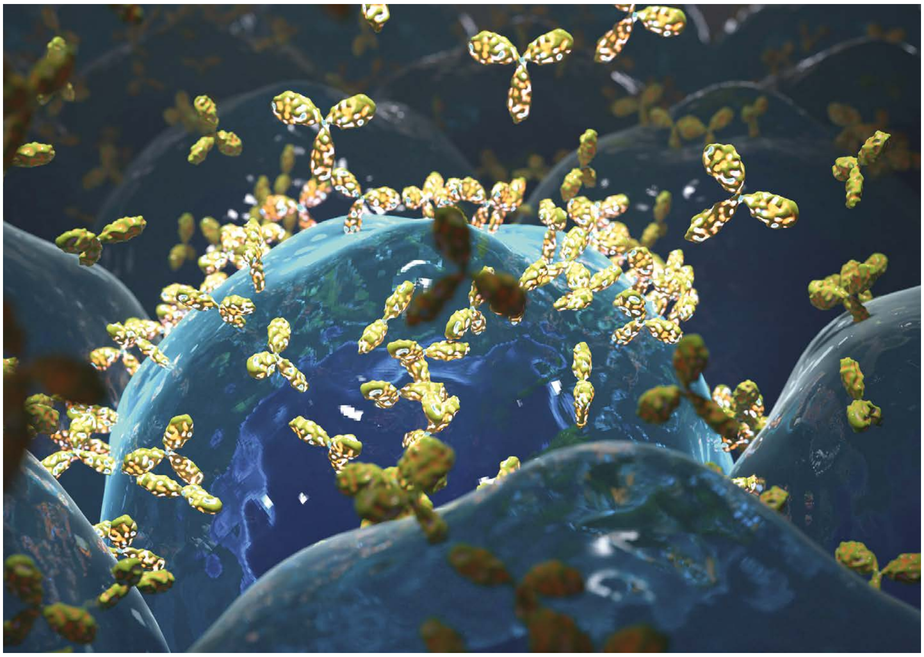
Beyond renaming

For many, institutional renaming is only a first step towards universities examining their own racist legacies and becoming more inclusive. Campus groups are now ratcheting up the pressure to diversify faculty and student bodies and to improve support for Black academics. "To me, the treatment of the people in the institutions matters just as much as the name that's on them," Lissanu says.

Jackson agrees that more action is needed. The renamings are "very low-activation-energy things", he says. "I'm happy they've done at least that." But he says he'd like to see policy changes with "far more teeth".

More transparency and accountability around how universities handle cases of racism would help to rebuild trust with Black academics, Jackson says. He also calls for universities to pay the students and faculty members who serve on diversity and equity committees. This sort of "invisible work" is important but isn't often rewarded monetarily or factored into career-advancement decisions.

Renaming buildings will be just a gesture if it is not backed up by meaningful change elsewhere on campus, says Ben Maldonado, who founded the Stanford Eugenics History Project, the student group that petitioned the university to rename Jordan Hall. And, he adds, that gesture is long overdue. "It's a thing you have to do but it's not something that you should praise Stanford – or anyone else – for doing."



NANOCLUSTERING/SPL

Molecules called monoclonal antibodies (artist's impression) could treat COVID-19.

CORONAVIRUS: WILL THE WORLD BENEFIT FROM ANTIBODY THERAPIES?

Monoclonal antibodies are expensive to produce, meaning poor countries might be priced out.

By Heidi Ledford

As the race to develop a vaccine against COVID-19 rages on, some researchers are focused on a short-term way to treat people with the disease: monoclonal antibodies. Rather than wait for vaccines to coax the body to make its own antibodies, these scientists want to inject designer versions of these molecules to directly disable the SARS-CoV-2 coronavirus. But mass-produced antibodies, routinely used to treat diseases such as cancer, are complex to manufacture and come with a hefty price tag. That risks placing them beyond the reach of poor countries.

That warning comes from a report released on 10 August by two leading charities: the International AIDS Vaccine Initiative (IAVI), a non-profit research organization in New York City, and Wellcome, a research funder in London. It calls for boosting the global availability of therapeutic antibodies against COVID-19 and other diseases by developing regulatory pathways, business models and technologies to lower the cost of the pricey medicine (see go.nature.com/30vwb5b).

It is a tall order, acknowledges Mark Feinberg, president of IAVI. "But COVID-19 really forces the issue in a major way," he says. "The pandemic demands that this dialogue take place."

Compelling science

A vaccine against COVID-19 is probably still months away, and it will be months after that before many people are able to receive it. Even then, some people, including older individuals, might not respond strongly to immunization, and others might refuse it altogether.

Those factors make it important to develop therapies against COVID-19. Physicians still don't have many ways to treat the disease. The antiviral drug remdesivir has been shown to shorten hospital stays for some patients, but it is expensive and in short supply. And a cheap steroid called dexamethasone has been shown to benefit only people with severe infections.

So scientists are increasingly focusing on monoclonal-antibody drugs in the hope that they will harness the immune system's natural response to viral invaders, says Jens Lundgren, an infectious-disease physician at the University of Copenhagen and Rigshospitalet, one of the city's hospitals. "The science around this has been exploding," he says. "It's very compelling." Lundgren is leading a large, multinational trial of an antibody developed by Eli Lilly in Indianapolis, Indiana; AbCellera in Vancouver, Canada; and the US National Institutes of Health (NIH).

In this approach, researchers isolate antibodies from recovering patients and identify those that best 'neutralize' the virus by binding to it and keeping it from replicating. They then produce these antibodies in bulk in the laboratory. If the treatment is found to be effective, companies will scale up production, using cells grown in giant bioreactors.

This differs from 'convalescent plasma' treatments, composed of a complex mixture of antibodies and molecules taken directly from the blood of people recovering from COVID-19 and used to treat other patients. The effects of both of these approaches are short term: neither type of treatment will produce a long-lasting immune response.

Access gap

IAVI estimates that more than 70 antibody therapies are being developed to treat and prevent COVID-19, and several clinical trials are under way.

But past experience suggests that if such treatments are developed against COVID-19, they might not find their way to much of the world. Monoclonal-antibody therapies are generally more expensive to make than are small-molecule drugs; they must be injected rather than taken orally; and they are difficult for generic-drug makers to duplicate. About 80% of global sales of licensed therapeutic antibodies – which treat autoimmune diseases, among other ailments – are in the United States, Europe and Canada. The median price for antibody therapies in the United States is US$15,000–200,000 per year of treatment, according to the IAVI–Wellcome report.

Feinberg says that the pandemic could spur technological innovation to find easier and cheaper ways to make large quantities of antibodies. It could also prompt business arrangements between the companies that develop therapeutic antibodies and other manufacturers – akin to the makers of generic versions of small-molecule drugs – that could try to copy the process and distribute the drugs more widely. And it might force regulators in low- and middle-income countries to become more familiar with antibody therapies and better able to approve their use.

"I don't know that any one of those will provide the solution," says Feinberg. "But if you combine them, then hopefully you'll have significant synergy."

Unique properties

No one has yet completed a large, randomized study of an antibody therapy against COVID-19, but results from such trials are expected in the coming months. Lundgren's trial, announced on 4 August, aims to enrol 1,000 people with COVID-19. Another large trial, sponsored by the NIH and Regeneron, a biotechnology company in Tarrytown, New York, launched on 6 July and will test a cocktail of two antibodies against SARS-CoV-2. Results are expected in late September.

Although these antibodies target the same virus, each interacts with SARS-CoV-2 differently: some will bind more strongly to the virus than will others, for example, or will target sites on its surface that shut the virus down more efficiently. And although antibodies are a natural means of defence, there are safety concerns, Lundgren notes. Researchers will be looking out for 'antibody-dependent enhancement', a phenomenon in which some antibodies can help viruses to gain entry into human cells, rather than prevent infection. A large trial is needed to settle the matter convincingly, Lundgren says.

OUTRAGE OVER RUSSIA'S FAST-TRACK CORONAVIRUS VACCINE

Scientists worry about the immunization's safety because it hasn't been tested in large trials.

By Ewen Callaway

Russian President Vladimir Putin announced on 11 August that the country's health regulator had become the first in the world to approve a coronavirus vaccine for widespread use — but scientists globally have condemned the decision as dangerously rushed. Russia hasn't completed large trials to test the vaccine's safety and efficacy, and rolling out an inadequately vetted vaccine could endanger people who receive it, researchers say. It could also impede global efforts to develop quality COVID-19 immunizations, they suggest.

"That the Russians may be skipping such measures and steps is what worries our community of vaccine scientists. If they get it wrong, it could undermine the entire global enterprise," says Peter Hotez, a vaccine scientist at Baylor College of Medicine in Houston, Texas.

"This is a reckless and foolish decision. Mass vaccination with an improperly tested vaccine is unethical. Any problem with the Russian vaccination campaign would be disastrous both through its negative effects on health, but also because it would further set back the acceptance of vaccines in the population," said Francois Balloux, a geneticist at University College London, in a statement distributed by the UK Science Media Centre.

Russian President Vladimir Putin receives a report about the coronavirus vaccine.

ALEXEI NIKOLSKY/SPUTNIK/EPA-EFE/SHUTTERSTOCK

In his announcement, Putin said that the Russian regulator had approved a COVID-19 vaccine developed by the Gamaleya Research Institute of Epidemiology and Microbiology in Moscow, even though phase III trials of the vaccine had yet to be completed. Such trials involve giving thousands of people a vaccine or a placebo injection, and then tracking them to see whether the vaccine prevents disease. The tests also allow researchers to confirm the vaccine's safety and look for rare side effects that might not have been observed in smaller, earlier-stage trials. Russian health-care minister Mikhail Murashko said at a government briefing that the vaccine would be gradually introduced to citizens, starting with health workers and teachers.

More than 200 COVID-19 vaccines are in development worldwide and several are already in phase III trials, with more front runners slated to begin theirs soon. But researchers think that even the earliest of those vaccines will not be approved for months.

Lack of data

The Gamaleya vaccine has been given to 76 volunteers as part of two early-stage trials listed on ClinicalTrials.gov, but no results from those trials or other preclinical studies have been published, and little else is known about the experimental vaccine.

According to the ClinicalTrials.gov listings, the vaccine, which is given in two doses, is made of two adenoviruses — viruses that cause a range of illnesses, including colds — that express the coronavirus's spike protein. The first dose contains an Ad26 virus — the same strain as is used in an experimental vaccine being developed by pharmaceutical company Johnson & Johnson of New Brunswick, New Jersey, and its subsidiary Janssen. The second, 'booster' dose is made of an Ad5 virus, similar to the one in an experimental jab being developed by CanSino Biologics in Tianjin, China.

According to the vaccine's Russian-language registration certificate, 38 participants who received one or two doses of the vaccine had produced antibodies against SARS-CoV-2's spike protein, including potent neutralizing antibodies that inactivate viral particles. These findings are similar to the results of early-stage trials of other candidate vaccines. Side effects were also similar, such as fever, headache and skin irritation at the site of injection.

Hotez expects that the Gamaleya vaccine will elicit a decent immune response against SARS-CoV-2. "The technical feat of developing a COVID-19 vaccine is not very complicated," he says. "The hard part is producing these vaccines under quality umbrellas — quality control and quality assurance — and then assuring the vaccines are safe and actually work to protect against COVID-19 in large phase III clinical trials."

But little is known about phase III trial plans for the Gamaleya vaccine. "I simply haven't managed to find any published details of a protocol," says Danny Altmann, an immunologist at Imperial College London. He hopes the trial is closely tracking the immune responses of participants and looking out for any side effects.

The head of a Russian government-owned investment fund said the vaccine would go through phase III testing in the United Arab Emirates, Saudi Arabia and other countries, according to the state-owned TASS Russian News Agency. The official said that purchase requests for one billion doses had been received from 20 countries in Latin America, the Middle East, Asia and elsewhere, and that manufacturing capacity was in place to produce 500 million doses, with plans for expansion.

'Ridiculous authorization'

Altmann is concerned that the vaccine could cause people who receive it and are then infected with SARS-CoV-2 to experience an exacerbated form of disease that occurs when antibodies generated by a vaccine carry the virus into cells. Another problem could be an asthma-like immune reaction that became an issue with some experimental vaccines against the related virus that causes SARS (severe acute respiratory syndrome). To spot these reactions, researchers would have to compare results from thousands of people who received a vaccine or placebo and then might have been exposed to SARS-CoV-2.

"It's ridiculous, of course, to get authorization on these data," says Svetlana Zavidova, head of Russia's Association of Clinical Trials Organizations in Moscow, which works with international pharmaceutical companies and research organizations. Without a completed phase III trial, Zavidova also worries that it will not be clear whether the vaccine prevents COVID-19 or not – and it will be difficult to tell whether it causes any harmful side effects, because of gaps in how Russia tracks the effects of medicines. "Our system for safety monitoring, I think, is not the best," she says.

Zavidova also worries the vaccine's approval will be "very harmful" for efforts to run clinical trials of other COVID-19 vaccines and other medicines in Russia.

"Not sure what Russia is up to, but I certainly would not take a vaccine that hasn't been tested in Phase III," tweeted Florian Krammer, a virologist at the Icahn School of Medicine at Mount Sinai in New York City. "Nobody knows if it's safe or if it works. They are putting [health-care workers] and their population at risk."

CONFERENCES FAILING TO PROTECT LGBT+ RESEARCHERS

Promoting equity, diversity and inclusion at meetings requires more than a code of conduct, analysis finds.

By Smriti Mallapaty

Ayesha Tulloch was reluctant to go to a conservation-biology conference in Malaysia, where laws discriminate against people of specific sexual orientations. "It came as quite a shock to me that the discipline I felt was the most accepting and tolerant toward the queer community would choose to have a conference in a place that's really not queer friendly," says Tulloch, a conservation scientist at the University of Sydney in Australia.

She did end up going to the meeting in Kuala Lumpur last year, organized by the Society for Conservation Biology (SCB), but she wondered whether the society's processes for fostering a diverse and inclusive meeting had failed when it chose that location.

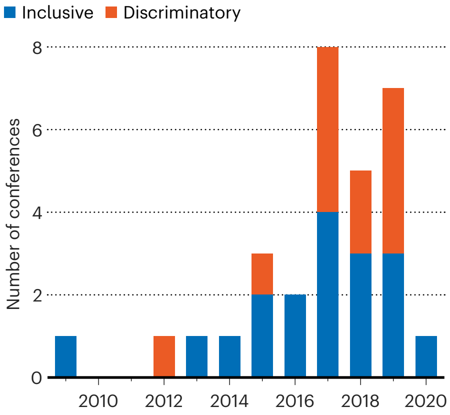
Tulloch went on to analyse policies and practices for supporting equity, diversity and inclusion around gender and sexual orientation, performing the first investigation of this kind. She looked at 30 ecology and conservation conferences held since 2009 and reported the results in *Nature Ecology and Evolution* on 3 August (A. I. T. Tulloch *Nature Ecol. Evol.* http://doi.org/d6nt; 2020). Tulloch found that about half of the events had codes of conduct promoting equity, diversity and inclusion. Those conferences were more likely than others to have initiatives that discouraged overt discrimination, such as a point of contact to report misconduct and facilities for breastfeeding and childcare.

No guarantee

But having a code did not always lead to initiatives that reduced implicit biases and barriers to participation, says Tulloch. For instance, conferences with a code were no more likely to advertise pronoun guidelines for name badges, select diverse speakers or choose locations safe for people of all genders and sexual orientations than were events without a code. Almost 40% of the conferences were held in locations where laws and societal norms discriminate against people of specific genders or sexual orientations (see 'Location, location'). And only two provided information on their websites about how they planned to ensure participants' general safety, for example by providing shuttle buses for safe transit between venues.

LOCATION, LOCATION

Some 40% of conservation and ecology conferences over the past decade were held in locations where laws and societal norms discriminate against people of specific genders or sexual orientations.



SOURCE: AYESHA TULLOCH

The analysis shows that codes of conduct have limitations, and putting a policy in place is not enough, says Lisa Kewley, an astrophysicist at the Australian National University in Canberra, who advocates for diversity at astronomy conferences.

But others say the analysis assumes that codes of conduct are supposed to promote diversity and inclusion, which is not necessarily their intended purpose. Codes are designed to protect against harassment and to clarify which behaviours will not be tolerated at a meeting, says Robyn Klein, a neuroimmunologist at Washington University in St. Louis, Missouri. They are not meant to have any bearing on diversity of speakers, she says.

Leslie Cornick, a conservation ecologist at the University of Washington Bothell who was chair of the 2019 SCB congress in Malaysia but had no part in deciding the location, agrees that codes of conduct are not necessarily intended to foster diversity, equity and inclusion, although they are a statement of values.

Cornick also notes that when choosing conference locations, organizers have to consider all members, including those who cannot afford to travel long distances.

But Tulloch says that codes are in place to address identity-based discrimination, which includes ensuring that participants have equal access. "The idea that a code is only there to prevent overt misconduct is outdated and incorrect," she says.

Q&A

LUKE PILKINTON-CHING/UNIV. OTAGO

New Zealand races to eliminate the coronavirus — again

At the start of this month, New Zealand was an exemplar for how swift and decisive action can stifle the spread of the coronavirus. No locally acquired cases of COVID-19 had been reported since the beginning of May. But the emergence of a cluster of cases — numbering 69 as *Nature* went to print — has caught the nation by surprise, and is a blow to the government's strategy to eliminate the virus. Amanda Kvalsvig, an epidemiologist at the University of Otago in Wellington, has been assisting with the country's COVID-19 response. She spoke to *Nature* about the rapid response to the new cases, and whether an elimination strategy is still possible.

How has the mood in New Zealand changed?

The new cases have been a shock. When they were announced, New Zealand had experienced more than 100 days with no identified community transmission, despite extensive testing. The country was at its lowest alert level, which allows near-normal activities, albeit with strict controls requiring travellers from overseas to remain in a quarantine facility for two weeks. There was a general feeling that we had beaten the virus — although government officials and public-health experts were warning against complacency.

Now, there's widespread anxiety, with long lines of people at COVID-19 testing stations and some people panic-buying in supermarkets.

What has been the public-health response to these new infections?

The response has been swift, backed up by decisive government action. The Auckland region, where the cases were identified, is now at Alert Level 3 — the second-highest of four levels — with people instructed to stay at home except for essential movement. The rest of the country is at Alert Level 2, which includes physical distancing measures and limits on mass gatherings.

People with COVID-19 are being tested, and their contacts traced. The government is now also recommending the use of face masks, and people with COVID-19 in the community will spend their isolation period in dedicated facilities instead of at home.

Population-wide mask use could help the country to avoid future lockdowns.

What is known about the original source of the outbreak?

The new cases came to light when a person in their fifties developed symptoms and presented for testing. Following that original positive test, their household and other contacts were tested, identifying further cases.

All of the new cases seem to be part of the same cluster, but that hasn't been linked back to its point of introduction into the country. That is concerning because we don't yet know how long this outbreak has been propagating. Ideally, investigations will allow the public-health system to 'backwards trace', identifying each source of the known cases, and then 'forwards trace' to identify other close contacts of that source.

Authorities are exploring the possibility that the virus arrived on packaging in cold storage. That's worth exploring, but global experience with COVID-19 outbreaks so far suggests that it is more likely to have originated from person-to-person close contact.

What could genomics tell us about this latest outbreak?

Genomic epidemiology is a powerful tool for tracing outbreaks back to the source, so it's particularly relevant to the current situation, where the original case is still unknown. If all of the Auckland cases turn out to be from one cluster, that will be good news for outbreak control. If there's more than one cluster, it will suggest more widespread transmission.

New Zealand has adopted an elimination strategy. Does this latest outbreak suggest that isn't possible?

We know that elimination is possible because New Zealand eliminated community transmission before. We expect to move in and out of elimination for the foreseeable future. The goal is to maintain zero community spread but this country will always be under threat from infections being introduced through the borders.

We've been fortunate to have outstanding political and scientific leadership. This has generated rapid and decisive action to protect population health. A key element of New Zealand's response has been excellent communication with the public about what is happening and what is expected of them.

Interview by Dyani Lewis

This interview has been edited for length and clarity.

Restrictions have been reintroduced in New Zealand after a new coronavirus outbreak.

LI QIAOQIAO/XINHUA/ZUMA PRESS

Feature

THE ANTIBIOTIC GAMBLE

Paratek Pharmaceuticals made a life-saving drug and got it approved. So why is the company's long-term survival still in question? **By Maryn McKenna**

HANNAH YOON FOR *NATURE*

Evan Loh, chief executive of the US firm Paratek Pharmaceuticals, leads a team that is striving to secure the future of a new antibiotic.

As the COVID-19 pandemic caught hold early this year, a small drug company outside Philadelphia was struggling to market a compound that could help patients battling for their lives.

Paratek Pharmaceuticals had spent more than 20 years developing and testing an antibiotic named omadacycline (Nuzyra), which went on sale in the United States in 2019 for use against bacterial infections. Although antibiotics can't fight the virus that causes COVID-19, almost 15% of people hospitalized with the disease go on to develop bacterial pneumonias, some of which are resistant to existing antibiotics.

Before COVID-19, antibiotic resistance was estimated to kill at least 700,000 people each year worldwide. That number could now climb as more people with the viral disease receive antibiotics to treat secondary infections, or to prevent infections that come from being on a ventilator. That's where a drug such as omadacycline might help – if it can be delivered to people in time to save lives.

"COVID is a wake-up call," says Evan Loh, chief executive of Paratek, which has offices in Pennsylvania and Boston, Massachusetts. Diagnostics, antibodies and vaccines are all key to preparing for a pandemic, he says, and "We need antibiotics, to give people the best chance of surviving this particular infection." But drug makers who produce antibiotics face unique challenges.

In a bitter paradox, antibiotics fuelled the growth of the twentieth century's most profitable pharmaceutical companies, and are one of society's most desperately needed classes of drug. Yet the market for them is broken. For almost two decades, the large corporations that once dominated antibiotic discovery have been fleeing the business, saying that the prices they can charge for these life-saving medicines are too low to support the cost of developing them. Most of the companies now working on antibiotics are small biotechnology firms, many of them running on credit, and many are failing.

In just the past two years, four such companies declared bankruptcy or put themselves up for sale, despite having survived the perilous, decade-long process of development and testing to get a new drug approved. When they collapsed, Achaogen, Aradigm, Melinta Therapeutics and Tetraphase Pharmaceuticals took out of circulation – or sharply reduced the availability of – 5 of the 15 antibiotics approved by the US Food and Drug Administration (FDA) since 2010 (see 'Trimming a thinning herd').

Paratek has so far avoided the rip tide that pulled so many others down, through a combination of conservative spending, experience and good fortune, including a lucrative government contract awarded late last year. But omadacycline's earnings, although steady, have not yet ensured Paratek's long-term survival.

"At the end of the day, Paratek is still going to have to sell a drug," says David Shlaes, a former pharmaceutical executive who is now an antibiotic-development consultant and author. "And it's not at all clear it's going to be able to sell as much as it needs to sell to make a profit."

Costly business

Bringing a new antibiotic to market represents a Herculean feat. Only about 14% of antibiotics and biologicals in phase I trials are likely to win approval, according to the World Health Organization. A team of economists estimated[1] in 2016 that the cost of getting from first recognition of an active drug molecule to FDA approval in the United States was US$1.4 billion, with millions more required for marketing and surveillance after approval. When companies such as Eli Lilly or Merck made antibiotics in the mid-twentieth century, those costs could be spread across their many divisions. And when, as used to happen, big companies bought smaller ones whose new drugs showed preclinical promise, the purchase price covered any debt the small companies had incurred.

> "INVENT A BAD ANTIBIOTIC, AND NO ONE WILL USE IT. INVENT A REALLY GOOD ANTIBIOTIC, AND REALLY NO ONE WILL USE IT."

Those business models no longer exist. The trio that runs Paratek knows this because all three are big-company veterans. Loh worked at Wyeth Pharmaceuticals in Philadelphia with Adam Woodrow, Paratek's president and chief commercial officer, and with Randy Brenner, chief development and regulatory officer, on the successful antibiotic tigecycline (Tygacil), which was approved in 2005. (Wyeth sold its antibiotic portfolio to Pfizer in 2009.)

"When you come from a big company to a small company, your focus becomes: 'How do I make sure this company survives?'" says Brenner, who previously also worked at Pfizer in New York City and at Shire in Lexington, Massachusetts (now a subsidiary of Takeda Pharmaceutical Company in Tokyo). "Bigger companies don't need to think like that. No matter what happens to a product, the company survives."

Tigecycline is based on tetracyclines, one of the earliest classes of antibiotic; they were first used in 1948, just six years after penicillin's debut. Over the years, successive generations of tetracyclines arrived on the market and were undermined by resistance. Tigecycline's structure incorporates tweaks that let it avoid those resistance mechanisms, but this comes at a cost: the drug can only be given intravenously.

This was a limitation. An intravenous drug would usually be given in hospitals and medical centres, making it both more expensive and less accessible to patients. So, as tigecycline was being developed, physician-researcher Stuart Levy – one of the giants of US antibiotic-resistance research, based at Tufts University in Boston – proposed formulating yet another tetracycline relative that could also be delivered in pill form. With that goal in mind, he co-founded Paratek in 1996 with Walter Gilbert, a molecular biologist at Harvard University in Cambridge, Massachusetts, who had won a share of the 1980 Nobel Prize in Chemistry.

In its early years, Paratek formed partnerships with larger companies – the German company Bayer, then Merck, then Novartis in Basel, Switzerland. But each deal dissolved as the corporations shifted focus or regulatory changes made omadacycline a bad financial bet. By 2012, when Loh was recruited, Paratek had accomplished phase I and II clinical trials of its compound, and had amassed abundant data on its safety – but it was running out of money. Loh cut the staff from about 34 people to 6, closing the research laboratory while the executive team scrounged for funds. For nine months, they went without salaries.

"I had an insolvency attorney on retainer for 18 months," he recalls. "I talked to him every week. Should I open the doors on Monday? Did I have enough cash to do that?"

In 2014, Paratek went public in a manoeuvre called a reverse merger, folding itself into a US company named Transcept Pharmaceuticals that was already listed on the NASDAQ stock exchange, but which had seen disappointing sales and was running with a skeleton crew. The deal earned Paratek $110 million, enabling it to launch omadacycline's phase III trials and begin a careful restaffing programme. In October 2018, the FDA approved the drug in oral and intravenous formulations against two conditions: complicated skin infections and community-acquired bacterial pneumonia. The 22-year journey was over – but the landscape into which omadacycline would launch was nevertheless still hazardous.

Loh, a cardiologist who had led transplant programmes at two academic medical centres before turning to the pharmaceutical industry, knew that the drug was needed. But he was aware it would not be easy.

“There’s nothing that happens in a hospital that can be successful if you don’t have an antibiotic,” he says. “You can’t have surgeries. You can’t have transplants. You can’t do anything. We have a product that we believe saves lives. Until we can make that successful for the long term, our mission is not done.”

Limited lifespan

Antibiotics present an enduring economic puzzle. These drugs changed the world. Yet despite their unique power, the free market doesn’t value them.

The reasons are complex. Start with the obvious: antibiotics kill bacteria, living things that are constantly adapting to threats against their survival. As soon as a new compound is used, pathogens start evolving strategies to foil the attack. That means an antibiotic’s useful life, and thus its earning potential, can be limited – a situation that doesn’t occur for most other drugs.

The duration of a new antibiotic’s lifespan wouldn’t be that important if a company could sell a lot of it quickly, but both structural and ethical barriers work against that (see ‘Long path to profitability’). Take the structural ones first. Relatively few patients have resistant infections that need treatment with new antibiotics, whereas most other drug categories are used to treat large numbers of people. The US Centers for Disease Control and Prevention estimates that there are 2.8 million resistant infections annually in the United States. For comparison, 7.4 million people in the United States take insulin to treat diabetes on a daily basis.

By one estimate, a new antibiotic needs to make at least $300 million in annual revenue to be sustainable[2]. Other researchers estimate[3] that the entire US market for new antibiotics that work against carbapenem-resistant Enterobacteriaceae – one of the most resistant and most stubborn classes of infection – is $289 million per year.

In other words, “there’s room in this marketplace for maybe one drug”, Shlaes says. “There’s not room for more than one drug if people want a return on their investment.”

Only a few of the companies now making antibiotics earn $100 million or more a year from them, according to analyses by the investment firm Needham in New York City. Most of the rest hover between $15 million and $50 million per year.

Then there are the ethical quandaries. Because any exposure of bacteria to an antibiotic risks the development of resistance, using that drug to treat one patient risks diluting its power to save others in the future. Thus, rules observed across health care, broadly called antibiotic stewardship, call for new antibiotics to be deployed slowly. That protects their reliability in the long term, but ruins their sales. For instance, in 2018, three new antibiotics – including the one made by recently bankrupt Achaogen – were used in only 35% of cases that would have qualified for them[4]. That was a win for stewardship, perhaps. It was a literal loss for the companies whose drugs would otherwise have been used.

“THERE’S ROOM IN THIS MARKETPLACE FOR MAYBE ONE DRUG.”

John Rex, a physician and long-time drug developer who is chief medical officer at the antifungals company F2G in Manchester, UK, and Vienna, sums up the paradox in this way: “Invent a bad antibiotic, and no one will use it. Invent a really good antibiotic, and really no one will use it.”

Into the abyss

The 100-person team that makes up Paratek approached the end of 2019 in an unsettled mood. They were staring into what Woodrow calls “the abyss of commercialization: this three-year period where you spend a tremendous amount of money before you get any traction in terms of real sales”. The antibiotic was selling steadily, but slowly – it was on track to earn $13 million that year. Meanwhile, Woodrow, Loh and Brenner had committed to doing post-approval studies and surveillance that they estimated would cost $70 million. And they had lost a guiding light: Levy, their co-founder, died in September 2019.

Then Christmas came early. The Biomedical Advanced Research and Development Authority (BARDA), a US federal agency, awarded Paratek a 5-year, $285-million contract to procure omadacycline for front-line troops who might be exposed to the bioweapon anthrax. (The purchase validated Levy’s early insight on the value of an oral drug: endangered troops could pop the pills and move on, rather than be tied to intravenous drips.)

On receiving the news, Loh felt like he could finally exhale. “This is a massive number – a gift,” he said not long afterwards. “It gives us time to gain traction.”

The BARDA money acted like a bridge across the chasms that other companies had fallen into. In a small way, it also demonstrated the potential of incentives for repairing the antibiotic market, which policymakers in the United States and Europe have been debating for several years. There are two types, referred to as push and pull. ‘Pushes’ propel new drug candidates from small companies through clinical trials and past approval. ‘Pulls’ aim to ease the financial crunch after approval, when companies must promote their drug without violating antibiotic stewardship.

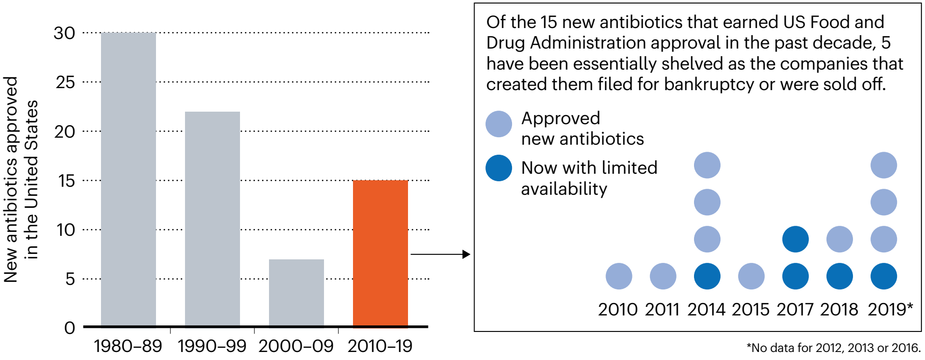
Push incentives have had some success. The non-profit organization CARB-X (Combating Antibiotic-Resistant Bacteria Biopharmaceutical Accelerator), based at Boston University, has gathered about $500 million in funding from US, UK and other European governments and philanthropies, and is distributing the money to small companies. Since CARB-X was founded in 2016, it has given 67 companies about $250 million to support promising preclinical and phase I research.

BARDA – which is funding the separate search for coronavirus vaccines and therapeutics – also gives push grants that support companies doing the later clinical trials that bring drugs to approval. However, BARDA’s contract with Paratek was different. It was effectively a pull incentive, an infusion of cash arriving after omadacycline had been approved, at a point when post-approval surveillance and studies to support use of the drug for other infections would eat up slender earnings.

Other forms of pull incentive have been proposed by analysts and lawmakers, among others, and considered by the US Congress, but they are much more controversial. These range from granting pharma companies extra time before other drugs they own become

TRIMMING A THINNING HERD

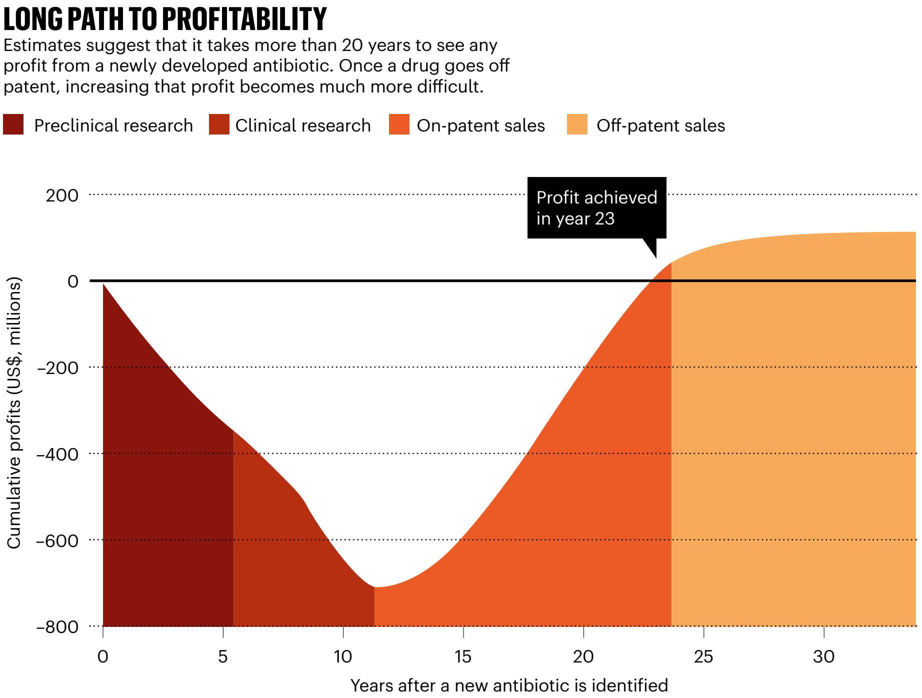
Over the past several decades, the number of new antibiotics approved for use in the United States has been declining, as it has elsewhere in the world.



Of the 15 new antibiotics that earned US Food and Drug Administration approval in the past decade, 5 have been essentially shelved as the companies that created them filed for bankruptcy or were sold off.

*No data for 2012, 2013 or 2016.

SOURCES: C. L. VENTOLA *PHARM. THER.* **40**, 277–283 (2015); AXIOS



LONG PATH TO PROFITABILITY

Estimates suggest that it takes more than 20 years to see any profit from a newly developed antibiotic. Once a drug goes off patent, increasing that profit becomes much more difficult.

Profit achieved in year 23

SOURCE: *SECURING NEW DRUGS FOR FUTURE GENERATIONS* (REVIEW ON ANTIMICROBIAL RESISTANCE, 2015)

generic, called extended market exclusivity, to giving companies market-entry rewards of billions of dollars that release them from the need to push sales of their drug, which would otherwise accelerate the development of resistance. Yet another proposed pull incentive – which would raise the reimbursements paid to hospitals by the US government for new antibiotics – was briefly added to the $2-trillion US stimulus bill written in response to the coronavirus pandemic. The incentive was taken out again before the bill became law.

No one has yet found a path past political reality: in the eyes of many voters and politicians, pharma companies are opportunists, inflating US drug prices to unconscionable heights. There were multiple congressional hearings on drug prices in 2019 alone, and in July, President Donald Trump signed several executive orders aimed at forcing prices down. Making things easier for any drug company, even a small one producing a much-needed antibiotic, faces strong political resistance.

Alan Carr, a molecular biochemist and senior analyst at Needham, says there is not yet a clear path to what works to support antibiotic research – not for incentives, and not for investors, either. "What has complicated things for investors is that there is a need for new antibiotics – but not in every space within antibiotics," he says. "There are certain infections where there's a real unmet need where we don't have any antibiotics. And then there are other areas where we have plenty. Unfortunately, what has happened is that investors have lumped the whole space together. So they want nothing to do with any of them."

Pandemic curveball

The BARDA contract turned Paratek from a company with less than a year's worth of cash in the bank to one that could count on funding to the end of 2023. That guaranteed its immediate future, although it did nothing to solve the long-term problem of needing to earn more from the drug than the market seemed willing to pay. And then the coronavirus hit.

When cases of SARS-CoV-2 started increasing in the United States, Loh and his team were unnerved. The Paratek sales force had been doing the normal rounds, explaining omadacycline to infectious-disease specialists and hospital pharmacists, hoping to have it picked up by the formulary committees that govern which medications hospitals routinely keep to hand. Its work was paying off. Month after month, sales of omadacycline were rising by more than 10%. When the lockdowns started, all of those meetings ended. The company worried its sales would stall as well. But in monthly data gathered since the epidemic began, the steady increase has continued.

"New prescribers, in a lockdown period – I expected that to go to zero," says Christine Coyne, Paratek's vice-president of marketing. "But we are still seeing double-digit growth."

It is too soon to say what drives those sales. Enough case reports have now been published[5,6] for researchers to feel confident that bacterial pneumonia is a complication of COVID-19 in 15–20% of patients. And in parts of the United States, the most common cause of bacterial pneumonia (*Streptococcus pneumoniae*) is resistant to azithromycin, the most common generic antibiotic, in up to 50% of cases. That could drive adoption of a new drug for which resistance has not been recorded. Other publications confirm that significant amounts of antibiotics are being prescribed to people with COVID-19 who are on ventilators, even when pneumonia has not been diagnosed (for a review, see ref. 7). This is an insurance policy against patients getting hospital-acquired infections, and because, in the absence of enough personal protective equipment, the procedures needed to confirm bacterial pneumonia are too risky for staff to undertake.

As a side effect of the pandemic, many other antibiotics are in short supply. That's a result of both interruptions in international trade – the active ingredients of most antibiotics come from China – and domestic influence. For instance, after Trump announced his support in March for the unproven and now largely discredited combination of hydroxychloroquine and azithromycin, several manufacturers of azithromycin announced that panic buying had triggered shortages.

If those events are boosting sales, that is to Paratek's benefit. They also underline the good fortune of the BARDA contract coming when it did. The company's supply chain avoids China and is based entirely in Europe. And, as a condition of protecting national defence, a clause in the BARDA contract requires the company to build a parallel supply chain fully within the United States, to avoid disruptions from any future outbreaks.

To the Paratek team, omadacycline's applicability to this ongoing crisis is validation of the company's commitment to stick with a product that it believed was needed. Equally, it has demonstrated how important it is to anticipate emergencies, and to provide for crucial medical interventions before one begins. The United States failed to do that for masks, respirators and other equipment that protects health-care workers from infection. It almost failed to do that for the provision of antibiotics, too.

"Coronavirus ought to say to the public, 'If you don't have technology on the shelf when something like this happens, you can't wait a year or two – or even three or five – in order to get it there,'" Loh says. "You can't be at the bedside and say to a company: 'Can you make this for me today?'"

Maryn McKenna is an independent journalist in Atlanta, Georgia, and a senior fellow of the Center for the Study of Human Health at Emory University in Atlanta.

1. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. *J. Health Econ.* **47**, 20–33 (2016).
2. Shlaes, D. M. *Antimicrob. Agents Chemother.* **64**, e02057-19 (2020).
3. Clancy, C. J. & Nguyen, M. H. *Antimicrob. Agents Chemother.* **63**, e01733-19 (2019).
4. Clancy, C. J., Potoski, B. A., Buehrle, D. & Nguyen, M. H. *Open Forum Infect. Dis.* **6**, ofz344 (2019).
5. Chen, N. *et al. Lancet* **395**, 507–513 (2020).
6. Zhou, F. *et al. Lancet* **395**, 1054–1062 (2020).
7. Rawson, T. A., Ming, D., Ahmad, R., Moore, L. S. P. & Holmes, A. H. *Nature Rev. Microbiol.* **18**, 409–410 (2020).

Books & arts

Neanderthal skeletons at the Smithsonian Museum of Natural History in Washington DC.

BILL O'LEARY/*THE WASHINGTON POST* VIA GETTY

No dullards, these Neanderthals

Horse eyeballs, shell tools and bone hammers – Rebecca Wragg Sykes paints a vivid portrait of our adaptable ancient relatives. **By Josie Glausiusz**

A quarter of the way through *Kindred*, I was longing to meet a Neanderthal. By the end, I realized that we had met. She is in me – or at least, in my genes.

In this deeply researched "twenty-first-century portrait of the Neanderthals" from birth to burial and beyond, palaeolithic archaeologist Rebecca Wragg Sykes smashes stereotypes. She ranges over 350,000 years, from the Neanderthals' first emergence more than 400,000 years ago to their disappearance about 40,000 years ago, describing how they bequeathed some of their genes to humans even as they vanished. Neanderthals were, she writes, "not dullard losers on a withered branch of the family tree, but enormously adaptable and even successful ancient relatives".

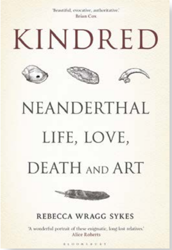
Based on fossil finds and artefacts from thousands of archaeological sites ranging from north Wales to the borders of China and the fringes of Arabia's deserts, hers are vivid, immersive depictions of Neanderthals from diverse periods and places. One imagines hunting with them, chewing on horse eyeballs, hammering stones into blades. And one pictures Neanderthals encountering our *Homo sapiens* ancestors, with whom they crossed paths and mated multiple times over a period of more than 100,000 years, as DNA evidence shows.

Distinct species

To conjure up this world, Wragg Sykes describes myriad discoveries, the first more than a century and a half ago. In the summer of 1856, limestone-quarry workers blasted open the Kleine Feldhofer Cave in the Neander Valley near Düsseldorf in what's now Germany, revealing ancient bones and the top of a skull. Scholars, including anatomist Hermann Schaaffhausen in Bonn, Germany, and geologist William King at Queen's College Galway in Ireland, speculated. Did the thick bones belong to a "barbarous and savage race" of humans (as proposed by Schaaffhausen)? Or had they come from an extremely ancient "pre-human"? It was King who named the species *Homo neanderthalensis*.

As further fossils were found – including the skeletons of two adults in Belgium in 1866 and a baby at the rock shelter of Le Moustier in France in 1914 – scholars agreed that Neanderthals were an extinct species distinct from humans. We now have specimens from between 200 and 300 Neanderthal individuals, ranging from newborns to adults in their fifties or even sixties, many just a single bone or jaw fragment.

And fossils tell only part of the story.



Kindred: Neanderthal Life, Love, Death and Art
Rebecca Wragg Sykes
Bloomsbury Sigma (2020)

"We have," Wragg Sykes notes, "millions more artefacts made by Neanderthals than bones from the hands that once touched them." Extensive studies of these finds have overturned old visions of fur-clad "brutes" (as King dubbed them) hunched over in the driving snow.

Take, for example, the period beginning around 130,000 years ago, called the interglacial Eemian. Average temperatures were 2–4 °C higher than today; melting polar caps and glaciers raised sea levels by some 8 metres; hippos and elephants inhabited what is now Britain. Europe's Neanderthals coexisted with Barbary macaques (*Macaca sylvanus*) – a species now confined to North Africa – as evidenced by fossil finds in the cave of Hunas in Germany. About 30 Neanderthal locales are known from this warm time. A 2018 study of 120,000-year-old lake-shore deposits at Neumark-Nord, Germany, shows that Neanderthals at the site used close-range thrusting spears to kill fallow deer (*Dama dama*; S. Gaudzinski-Windheuser *et al. Nature Ecol. Evol.* **2**, 1087–1092; 2018).

Visceral impulse

Neanderthals adapted with growing inventiveness to dramatic shifts in climate. "More artisans than klutzes," Wragg Sykes writes, they crafted dozens of types of stone blade, as well as long, finely tapered wooden spears, shell tools and bone hammers. They used tactical planning to ambush groups of prey, including bison, horses, rhinoceroses and reindeer.

Patterns of cut marks on skeletons show that Neanderthals favoured fat-rich brains, "as well as other juicy parts like eyeballs, tongues and viscera", and prized marrow-filled bones. They ate rabbits, birds and fish, gorged on tortoises and slaughtered hibernating bears. Analysis of charred hearths in the Kebara Cave in Israel and elsewhere show that Neanderthals also nibbled on nuts such as acorns and walnuts, and ate fruits including dates and figs, as well as wild radishes, peas and lentils.

LATITUDESTOCK/ALAMY

Researchers excavate Gorham's Cave in Gibraltar, where Neanderthals lived for 100,000 years.

Did they use language or engage in abstract thought? "Musing about minds from 50 or 100 millennia ago is of course fraught with pitfalls," Wragg Sykes cautions. Neanderthals had flatter foreheads than humans, with less space for the frontal cortex – which is intimately connected to memory and language. But computer modelling suggests that their vocal cords could make a range of sounds similar to ours, she says. In apparent artistic impulses, Neanderthals in many places used red-ochre pigments and might have ornamented themselves with feathers. One group engraved a cross-hatched grid pattern on the floor of Gorham's Cave, Gibraltar. Among their more mysterious creations are two rings of snapped-off stalagmites, arranged on the chamber floor of a cave near the French village of Bruniquel, dating to about 177,000 years ago.

> **"Neanderthals adapted with growing inventiveness to dramatic shifts in climate."**

Above all, Neanderthals were wanderers, Wragg Sykes shows. They were top-level hunters and foragers, and there were few landscapes they did not traverse . Their sites were not destinations but intersections, "nodes within networks stretching hundreds of kilometres". This nomadic tradition might have saved them from rising sea levels during the Eemian.

Sadly, a similar escape might not be an option for us, their human relatives. In her epilogue, written from home lockdown in early 2020 during the COVID-19 pandemic, Wragg Sykes warns that "we are heading into a world hotter and more dangerous than any previous hominin survived". As polar ice caps are at risk of disappearing, the Arctic, Amazon and Australia burn and heat records break like waves. She writes: "Eurasia with maybe a few hundred thousand souls is very different to today's teeming millions ... We have no guidebook for the destination our sprawling, industrialised, unimaginably complicated civilisation faces."

Josie Glausiusz is a science journalist in Israel.
Twitter: @josiegz

Readers respond

Correspondence

SDGs: affordable and more essential now

Your call to scale back the ambitions of the Sustainable Development Goals (SDGs; see *Nature* **583**, 331–332; 2020) conflates two issues. The first is whether the goals are technically and financially feasible. The second is whether they are likely to be accomplished under current policies.

The SDGs are, in principle, still affordable and achievable. But they are being undermined by the chronic failure of the United States and other rich nations to honour the goal of international partnership (SDG 17), as well as by failures in international cooperation and domestic governance of many countries.

Criticisms have not demonstrated any technological or operational obstacles to achieving the SDGs. Academic studies, commission reports and policy analyses suggest that there are pathways to success in areas such as energy decarbonization, sustainable land use and food systems, education for all, disease control and public health. They rely on a combination of policies, including transfers of public funds to poor people, public financing of health care and education, and increased public and private investment in infrastructure.

The goals are affordable. Assessments by the International Monetary Fund, the United Nations Sustainable Development Solutions Network and others confirm that the SDGs can be financed at a cost of about 2% of global gross domestic product, with around 0.4% in development aid to fill the gaps in lower-income countries. Ambitious goals unleash innovations to accelerate progress and bring down costs, particularly through the use of new technologies.

In this way, ambitious goals have helped to achieve tremendous advances in the control of infectious diseases that many experts had considered impossible (J. D. Sachs and G. Schmidt-Traub *Science* **356**, 32–33; 2017). However, most rich nations do not spend the minimum target of 0.7% of their gross national income on 'official development assistance'.

The COVID-19 pandemic is a serious setback for sustainable development. Had the SDGs been heeded sooner, control today would be faster and more effective. SDG 3.d calls for "early warning, risk reduction and management of national and global health risks", which many countries, including wealthy ones, have overlooked. The SDGs provide an inclusive framework for post-COVID-19 economic recovery, and for development decoupled from negative environmental impacts (http://sdgindex.org/).

Rather than abandoning goals that reflect basic human rights and ignoring the need to respect Earth's planetary boundaries, experts should uphold the SDGs and speak truth to power about what is needed to achieve them.

Jeffrey Sachs, **Guido Schmidt-Traub**, **Guillaume Lafortune** UN Sustainable Development Solutions Network, Paris, France.
guido.schmidt-traub@unsdsn.org

Maasai teacher Isaac Mkalia consults his mobile phone in Kenya.

SVEN TORFINN/PANOS

SDGs: aggregate to fix prioritization

The COVID-19 pandemic hinders achievement of some of the United Nations Sustainable Development Goals (SDGs; see *Nature* **583**, 331–332; 2020), but it has revealed the greater importance of those related to health and safety. I agree that considering them equally important might be unrealistic (R. Naidoo and B. Fisher *Nature* **583**, 198–201; 2020). An aggregated approach would allow for trade-offs between and prioritization of different goals.

Existing frameworks for a single outcome — such as normalizing scores across countries — can be simplistic and lack ethical underpinnings (T. Schaubroeck *et al. Environ. Sci. Technol.* **54**, 2051–2053; 2020). A better way to assess sustainable development, dealing with human needs, would be to use well-being as the end goal.

The original SDGs could be complemented by a flexible aggregated approach that can be applied differently in various scenarios, such as lockdowns versus no lockdowns.

Thomas Schaubroeck Luxembourg Institute of Science and Technology, Belvaux, Luxembourg.
thomas.schaubroeck@list.lu

SDGs: a North Star to guide us through this dark time

In a multipronged global crisis, now is not the time to reconsider the United Nations Sustainable Development Goals (SDGs; *Nature* **583**, 331–332; 2020). The COVID-19 crisis stems from exactly the type of interconnected failure that the SDGs aim to address. This moment requires absolute clarity while we continue to fight for the world that we need.

Although many SDGs might now seem harder to achieve, the pandemic is not a reason to scale them back. On the contrary, it reinforces why the goals were established in the first place: to chart a better course towards common economic, social and environmental ambitions that will guarantee humanity's long-term future. COVID-19 does not alter the need to reduce greenhouse-gas emissions or ocean acidification. Nor does it mitigate the need to end pointless deaths and persistent inequities.

In 2015, the SDGs emerged from a painstaking 3-year diplomatic negotiation among 193 countries. Amid current geopolitical tensions, it is unlikely that all these countries could reach a better consensus today — on this or any topic. Whatever their imperfections, the SDGs are a 'North Star' to help us to rebuild after today's crisis.

Amar Bhattacharya, **Homi Kharas**, **John W. McArthur***
Brookings Institution, Washington DC, USA.
jmcarthur@brookings.edu

---

*Declares non-financial competing interests; see go.nature.com/2xvgy0x

Expert insight into current research

News & views

Coronavirus

COVID-19 poses a riddle for the immune system

Stanley Perlman

It is unclear why people's immune response to the SARS-CoV-2 coronavirus varies so widely. Tracking patient responses over time sheds light on this issue, and has implications for efforts to predict disease severity. **See p.463**

A dysregulated immune response, a cytokine storm and cytokine-release syndrome[1,2] are some of the terms used to describe the over-exuberant defence response that is thought to contribute to disease severity in certain people who become seriously ill with COVID-19. However, a precise definition of this type of immune dysfunction remains elusive. On page 463, Lucas *et al.*[3] fill in some gaps in our knowledge.

A holy grail of COVID-19 research is the ability to assess a person's immune response, to pinpoint early the individuals who have mild symptoms but who are on track to develop the intense defence response that is associated with severe disease. This is important because there is a broad spectrum of clinical disease in people infected with SARS-CoV-2, the coronavirus that causes COVID-19: some infected individuals can be asymptomatic, whereas others are at risk of dying, and require hospitalization in an intensive-care unit and use of a ventilator machine to breathe[4,5]. Identifying those whose dysregulated immune-response signature predicts the development of severe disease would enable them to be monitored more intensively to minimize disease progression.

Lucas and colleagues performed extensive analyses of immune responses over time (longitudinal studies) in 113 people hospitalized with COVID-19 who had moderate or severe disease, and assessed a similar number of SARS-CoV-2-free healthy people as controls. The authors analysed molecules in blood plasma (Fig. 1) and monitored peripheral blood mononuclear cells – white blood cells of the immune system such as CD4 T cells, CD8 T cells and B cells. The longitudinal nature of this study enables conclusions to be drawn that wouldn't be possible from analysing cross-sectional studies that don't follow individuals over time.

The authors found that levels of several molecules that promote inflammation – immunomodulatory molecules termed cytokines, including IL-1α, IL-1β, IFN-α, IL-17A and IL-12 p70 – were higher in all of the people who had COVID-19 than in the healthy controls, providing a 'core' COVID-19 signature. Other cytokines, such as IFN-λ, thrombopoietin (which is associated with abnormalities in blood clotting), IL-21, IL-23 and IL-33, were upregulated to a greater extent in people with severe COVID-19 than in those with moderate disease. Several of the molecules upregulated in the core COVID-19 signature, as well as those seen in severe disease, have been identified previously as positively correlated with COVID-19 severity[6,7]. Severe disease was characterized by prolonged elevation of many of these molecules, whereas the levels of most of them subsided in people with moderate disease. Moreover, individuals with severe disease showed increased levels of cytokines associated with activation of a protein complex called the inflammasome, a component of the immune response that is a driver of inflammation. Also increased were levels of IL-1Ra, a protein that normally inhibits excessive inflammasome function, providing a rare example of an upregulated molecule that dampens the immune response in severe disease.

Levels of molecules associated with a defence response to viral infection – released by a type of activated CD4 T cell called a

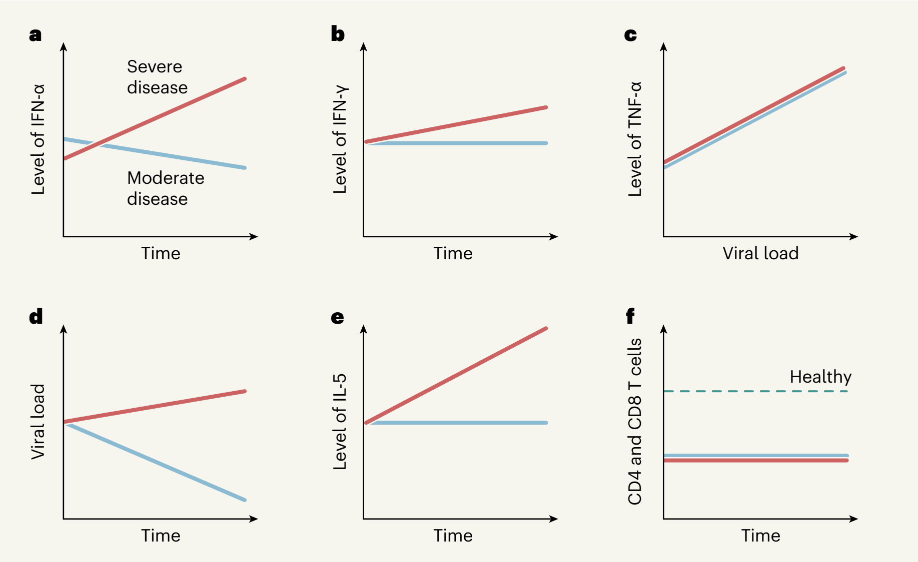


Figure 1 | Immune responses to COVID-19 infection. Lucas *et al.*[3] analysed blood samples taken over time from individuals hospitalized with moderate or severe COVID-19. Such information is useful for efforts to try to predict the individuals at risk of developing a severe form of the disease, which is often accompanied by an intense immune response. **a**, The authors identified a subset of immune-signalling molecules called cytokines that are expressed in people with moderate or severe disease; IFN-α is one such 'core' cytokine. **b**, The expression level of certain other cytokines, such as IFN-λ, mainly changed when the disease became more severe. **c**, The level of some inflammation-promoting cytokines, such as TNF-α, correlated with viral load in the nasal passages. **d**, Viral load declined over time in people with moderate COVID-19, but not in those with severe disease. **e**, Some cytokines not associated with antiviral responses, such as IL-5, which aids defence against parasitic worms and is released during allergic reactions, were, surprisingly, upregulated as people developed severe disease. **f**, The levels of CD4 and CD8 T cells, which are key immune cells involved in viral clearance, were lower in people with moderate or severe disease than in healthy individuals uninfected with SARS-CoV-2, the virus that causes COVID-19. (Graphs based on data from ref. 3.)

T_H1 cell — were higher in people with severe disease than in those with moderate COVID-19. This occurred even though blood levels of CD4 T cells and CD8 T cells, which are generally linked to expression of these molecules, were similarly decreased (a condition called lymphopenia) in people with moderate or severe disease. More remarkably, cytokines associated with immune responses to fungi (cytokines released by a type of CD4 T cell called a T_H17 cell) were elevated and remained so in people with severe disease. The same was true for cytokines associated with immune responses to parasites, including worms, or with allergic reactions (cytokines such as IL-5, released by a type of CD4 T cell called a T_H2 cell). The discovery that parts of the immune system unrelated to viral control would be triggered by a viral infection was unexpected. Less surprising was the finding that levels of inflammatory cytokines in the blood, especially the proteins IFN-α, IFN-γ, TNF-α and TRAIL, correlated with viral RNA levels in the nasal passage, independently of disease severity.

From their analysis of proteins in people's peripheral blood mononuclear cells, the authors divided individuals into three groups on the basis of their subsequent clinical course and disease severity. In general, at early time points after infection, those who went on to have moderate disease had low levels of inflammatory markers and a rise in the level of proteins associated with tissue repair. By contrast, people who went on to develop severe or very severe disease had increased expression of IFN-α, IL-1Ra and proteins associated with T_H1-, T_H2- and T_H17-cell responses, even at early time points (10–15 days after the onset of symptoms). These results were validated using data for the entire patient population, across all time points, thus demonstrating that these characteristic expression patterns persisted over time in people with each type of disease severity.

What have we learnt from this report, and what still needs to be done? It is clear from this and other studies that the immune response in hospitalized patients with severe COVID-19 is characterized by lymphopenia and the expression of molecules associated with ongoing inflammation[8], whereas these same molecules are expressed at a lower level in people with mild or moderate disease. Differences in immune responses between the different categories of disease severity are even more evident when people with very mild or subclinical disease are included in the analyses[4].

A key next step will be to analyse samples from people with extremely early signs of COVID-19, and to compare longitudinal data in those who do and those who don't require hospitalization. Some people who develop severe disease seem to have a suboptimal immune response initially, which might allow uncontrolled viral replication[9]. Such high replication might, in turn, contribute to severe disease.

Further analyses should identify molecules that are useful for predicting which individuals will later be hospitalized and require intensive care. It will also be crucial to understand how severe disease results in an upregulation of cytokines usually linked to the immune response to parasites and allergic reactions, and whether this apparent dysregulation of the immune response to viral infection is unique to COVID-19. It will also be worth determining whether these changes in the expression of inflammatory molecules in the blood also occur in cells at the site of infection — the airways and lungs. Lucas *et al.* analysed blood samples because obtaining cells from an infected lung is much more tricky and results in the production of aerosols that might contain SARS-CoV-2.

For results to be clinically useful, it will be necessary to define a limited number of biomarkers that can be both readily measured and used to predict disease outcomes. This could be difficult, because many of the changes in cytokine expression observed in studies such as that of Lucas and colleagues are useful for population-level analyses but less so for predicting outcomes in individual patients. Levels of specific cytokines vary substantially between people, making it hard to benchmark a level of cytokine expression that constitutes a sign of abnormality. Therefore, groups of cytokines, each with different degrees of inter-individual variability, must be measured to identify useful alterations.

The identification of infected people on course to develop severe COVID-19 will be a key step forward in patient care. For example, it would increase the possibility of correctly selecting individuals most in need of targeted early treatment, such as with therapies that directly inhibit viral replication. There has been progress in identifying such treatments, and the continued development of antiviral drugs that have increased efficacy and specificity will be crucial for alleviating the disease and reducing the death rate associated with the COVID-19 pandemic. Ideally, such drugs will be administered orally, and will reduce the need for hospitalization. Continued progress in unravelling the immune response to SARS-CoV-2 infection will help to improve clinical treatments for COVID-19.

Stanley Perlman is in the Department of Microbiology and Immunology, University of Iowa, Iowa City, Iowa 52242, USA.
e-mail: stanley-perlman@uiowa.edu

1. Moore, J. B. & June, C. H. *Science* **368**, 473–474 (2020).
2. Hirano, T. & Murakami, M. *Immunity* **52**, 731–733 (2020).
3. Lucas, C. *et al. Nature* **584**, 463–469 (2020).
4. Long, Q.-X. *et al. Nature Med.* https://doi.org/10.1038/s41591-020-0965-6 (2020).
5. Wang, D. *et al. J. Am. Med. Assoc.* **323**, 1061–1069 (2020).
6. Zhou, Z. *et al. Cell Host Microbe* **27**, 883–890 (2020).
7. Lee, J. S. *et al. Sci. Immunol.* **5**, eabd1554 (2020).
8. Zhang, X. *et al. Nature* **583**, 437–440 (2020).
9. Mathew, D. *et al. Science* https://doi.org/10.1126/science.abc8511 (2020).

Ecology

Species that can make us ill thrive in human habitats

Richard S. Ostfeld & Felicia Keesing

Does the conversion of natural habitats to human use favour animals that harbour agents causing human disease? A global analysis of vertebrates provides an answer to this pressing question. **See p.398**

Humans have altered more than half of Earth's habitable land to meet the needs of our burgeoning population[1]. The transformation of forests, grasslands and deserts into cities, suburbs and agricultural land has caused many species to decline or disappear, whereas others have thrived[2]. The losers tend to be ecological specialists, such as rhinoceros or ostriches, that have highly specific feeding or habitat requirements and that are comparatively larger, rarer and longer-lived than are non-specialists. The winners are often generalists that are small and abundant and that have 'fast', short lives, such as rats and starlings.

On page 398, Gibb *et al.*[3] show that, worldwide, these winners are much more likely to harbour disease-causing agents (pathogens) than are the losers. As a result, when we convert natural habitats to our own uses, we inadvertently increase the probability of transmission of zoonotic infectious diseases, which are caused by pathogens that can jump from animals to humans.

Examples of how land-use change increases the risk of zoonotic disease have been accumulating for decades. For example, rodents that amplify the abundance of pathogens that cause Chagas disease, several tick-borne illnesses and a suite of what are termed hantaviral diseases thrive in human-dominated landscapes where other species have been lost[4]. But the generality of this pattern, and the specific mechanisms that underlie it, have been questioned[5].

Gibb and colleagues had to overcome two obstacles in investigating whether, at a global scale, human-caused changes to ecosystems favour vertebrate species that are most likely to cause illness. One challenge was determining which animal species tend to disappear and which tend to thrive, along a gradient from undisturbed, natural habitats to the most human-dominated areas. The authors accomplished this using the database of the PREDICTS project (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems). It contains more than 3.2 million records from 666 studies that counted animals along land-use gradients around the world[6].

The second hurdle was determining which of these species harbour pathogens that can infect humans. To do this, Gibb *et al.* compiled information from six databases that report host–pathogen associations. They found 20,382 associations between 3,883 vertebrate host species and 5,694 pathogens. Unfortunately, finding that an animal and a pathogen are associated does not necessarily indicate that the animal can transmit the pathogen to humans or other animals. Recognizing this, Gibb and colleagues used more-stringent criteria to ascertain host–pathogen associations, including determining whether there was direct evidence of the pathogen existing in the host, and of the host's ability to transmit the pathogen.

The patterns that the authors detected from these analyses were striking. As human-dominated land use increased, so did the total number of zoonotic hosts, whereas the total number of non-hosts declined. In more intensively used areas, both the number of host species and the number of individuals of those species increased, with the latter effect being the stronger of the two. The abundances of rodents, bats and songbirds increased notably in human-dominated sites (Fig. 1). The effect on the abundances of carnivores and primates was more modest. However, host species could be misclassified as non-host species if a lack of in-depth research effort resulted in a failure to detect zoonotic pathogens. To take this into account, Gibb *et al.* incorporated a statistical process called bootstrapping into their analysis. This allowed them to reclassify non-hosts to host status using an approach that included the amount of published research on the species. Their conclusions using this approach remained the same.

ALAMY

Figure 1 | A rat on a city street. Gibb *et al.*[3] report that vertebrates, such as rodents, that can harbour agents that cause human disease flourish in human-altered landscapes.

The COVID-19 pandemic triggered by a coronavirus of animal origin has awakened the world to the threat that zoonotic diseases pose to humans. With this recognition has come a widespread misperception that wild nature is the greatest source of zoonotic disease. This idea is reinforced by popular-culture portrayals of jungles teeming with microbial menaces, and by some earlier scientific studies[7,8]. Gibb *et al.* offer an important correction: the greatest zoonotic threats arise where natural areas have been converted to croplands, pastures and urban areas.

Is it simply a coincidence that the species that thrive in human-dominated landscapes are often those that pose zoonotic threats, whereas species that decline or disappear tend to be harmless? Is the ability of animals to be resilient to human disturbances linked to their ability to host zoonotic pathogens? Gibb *et al.* found that the animals that increase in number as a result of human land use are not only more likely to be pathogen hosts, but also more likely to harbour a greater number of pathogen species, including a greater number of pathogens that can infect humans.

Using a different approach to address the same general questions, a recent study[9] found that mammals that are increasingly widespread and abundant carry more zoonotic viruses than do mammals that are declining, threatened or endangered. These observations support previous research that documents a trade-off between the high reproductive rates associated with ecological resilience and the high immune-system investment associated with lower pathogen loads[10]. In other words, creatures that have rat-like life histories seem to be more tolerant of infections than do other creatures. An alternative, although not mutually exclusive, explanation is that generalist pathogens, which are more likely to spill over into new hosts, tend to adapt to target the hosts they are most likely to encounter over evolutionary time[11]. These hosts are the rats, and not the rhinos, of the world.

The analyses by Gibb *et al.* and others[9] suggest that restoring degraded habitat and protecting undisturbed natural areas would benefit both public health and the environment. And, going forward, surveillance for known and potential zoonotic pathogens will probably be most fruitful if it is focused on human-dominated landscapes.

Richard S. Ostfeld is at the Cary Institute of Ecosystem Studies, Millbrook, New York 12545, USA. **Felicia Keesing** is in the Biology Program, Bard College, Annandale-on-Hudson, New York 12504, USA.
e-mails: ostfeldr@caryinstitute.org; keesing@bard.edu

1. Field, C. *et al.* in *Planetary Health: Protecting Nature to Protect Ourselves* (eds Myers, S. & Frumkin, H.) 71–110 (Island, 2020).
2. Newbold, T. *et al. Nature* **520**, 45–50 (2015).
3. Gibb, R. *et al. Nature* **584**, 398–402 (2020).
4. Ostfeld, R. S. & Keesing, F. *Annu. Rev. Ecol. Evol. Syst.* **43**, 157–182 (2012).
5. Rohr, J. R. *et al. Nature Ecol. Evol.* **4**, 24–33 (2020).
6. Hudson, L. N. *et al. Ecol. Evol.* **7**, 145–188 (2017).
7. Jones, K. E. *et al. Nature* **451**, 990–993 (2008).
8. Allen, T. *et al. Nature Commun.* **8**, 1124 (2017).
9. Johnson, C. K. *et al. Proc. R. Soc. B* **287**, 20192736 (2020).
10. Previtali, M. A. *et al. Oikos* **121**, 1483–1492 (2012).
11. Ostfeld, R. S. & Keesing, F. *Can. J. Zool.* **78**, 2061–2078 (2000).

Structural biology

How lipopolysaccharide strikes a balance

Russell E. Bishop

Bacteria with two membranes must regulate the production of a surface molecule known as lipopolysaccharide. The structure of an essential signal-transduction protein now reveals how lipopolysaccharide controls its own synthesis. **See p.479**

Feedback inhibition occurs when the product of a metabolic pathway diminishes its own production by triggering a decrease in the activity of a key enzyme in the pathway. Such inhibition controls the production of lipopolysaccharide (LPS) molecules, which are an integral part of the outer membrane of some bacteria. It has long been suspected that the feedback signal responsible for regulating LPS biosynthesis is either LPS itself, or one of its precursors[1]. But, on page 479, Clairfeuille *et al.*[2] add to a flurry of recent work[3–5] showing that the membrane protein PbgA is the long-sought LPS signal transducer in the bacterium *Escherichia coli*. The current study extends our understanding of PbgA by providing a high-resolution structure of the protein bound to LPS.

E. coli has two distinct membranes: the inner membrane, which is a phospholipid bilayer; and the asymmetric outer membrane, in which LPS lines the external surface, and a single layer of phospholipids forms the internal surface. LPS provides a barrier to greasy antibiotics and detergents that are encountered in the gut of mammalian hosts. The ratio of phospholipid to LPS is crucial for membrane function — too much LPS is toxic to the inner membrane and too little compromises the outer membrane (reviewed in ref. 1).

LPS assembly starts on the internal surface of *E. coli*'s inner membrane. The rate of assembly is controlled by the enzyme LpxC. Before LPS generation is completed, the lipid is flipped to the external surface of the inner membrane for further modification. The completed LPS is then transported to the external surface of the outer membrane by means of a protein bridge that connects the membranes (reviewed in ref. 1).

Investigations[3–5] published this year of how this pathway is regulated have produced a model in which PbgA on the inner membrane modulates the activity of LpxC by interacting with LapB — a protein that guides the enzyme FtsH to degrade LpxC (ref. 1). So when levels of LPS are low, PbgA inhibits the interaction between LapB and FtsH in the inner membrane, stabilizing LpxC and promoting LPS biosynthesis (Fig. 1a). When the number of LPS molecules exceeds a threshold in the outer membrane, LPS transport across the bridge ceases[6]. LPS accumulates on the external surface of the inner membrane, which can cause the formation of potentially lethal irregular membrane structures[3]. By sensing the accumulated LPS, PbgA can relax its inhibition of LapB–FtsH. LpxC can be degraded, thus diminishing LPS biosynthesis and restoring the phospholipid–LPS balance (Fig. 1b). Clairfeuille and colleagues' work now points to the same mechanism for LPS sensing, adding weight to this emerging model.

The authors corroborated the finding[3–5] that *E. coli* strains carrying truncated forms of PbgA (which lack extracellular and linker domains that normally connect to its essential transmembrane domain) remain viable, but are chronically deficient in LPS (Fig. 1c). In these mutants, phospholipids migrate into the external surface of the outer membrane to create mixed membranes containing patches of phospholipid bilayer scattered among the zones of LPS–phospholipid membrane. The phospholipid bilayer patches allow greasy antibiotics and detergents to enter the cell, and transient defects at the boundaries between the two different lipid phases allow leakage of large soluble molecules[7].

Previous work has shown that a greasy functional group called palmitate is incorporated into LPS when phospholipids are present at the external surface[8]. Clairfeuille *et al.* demonstrate the presence of palmitate in the outer-membrane LPS of a PbgA mutant.

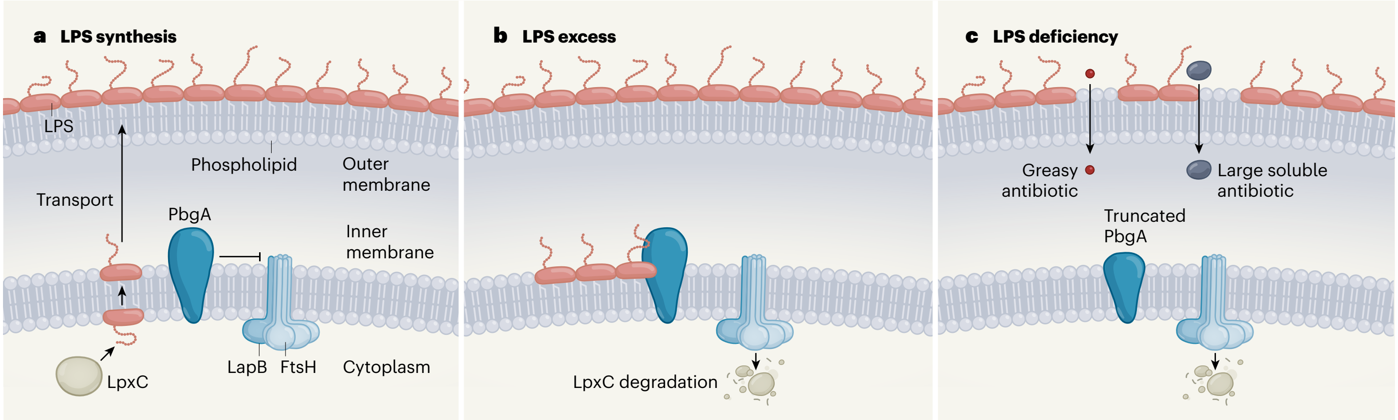


Figure 1 | Feedback inhibition regulates lipopolysaccharide biosynthesis. The bacterium *Escherichia coli* has an inner membrane comprising two phospholipid layers and an outer membrane, which has one layer of phospholipids and one layer of lipopolysaccharide (LPS) molecules. **a**, The enzyme LpxC controls the biosynthesis of LPS from precursors in the cell cytoplasm. After being flipped to the external surface of the inner membrane, the mature LPS is then transported to the outer membrane. The FtsH enzyme, guided by interactions with LapB protein, degrades LpxC — but Clairfeuille *et al.*[2] and others[3–5] show that the protein PbgA inhibits LapB–FtsH activity, and so promotes LPS biosynthesis. **b**, When excess LPS accumulates on the external surface of the inner membrane, it binds to PbgA. The protein relaxes its control on LapB–FtsH, allowing degradation of LpxC to restore normal LPS levels. **c**, A PbgA truncation mutation leads to chronic depletion of LPS, presumably because the mutant only weakly inhibits LapB–FtsH. Phospholipids fill the gaps left by LPS in the outer membrane, enabling greasy antibiotics and detergents to penetrate at local phospholipid bilayers, and large soluble compounds to leak through transient boundary defects where the LPS and phospholipid phases meet.

Such palmitate incorporation has also been reported in bacteria carrying mutations in components of the transport systems that move LPS towards the outer membrane[9] and phospholipids away from it[10,11]. What can these observations tell us about the function of PbgA? They could fit with the proposal[12,13] that PbgA is a transport protein for the phospholipid cardiolipin. However, directly blocking LPS biosynthesis can also lead to LPS depletion, and to incorporation of palmitate in outer-membrane LPS[14,15]. As such, PbgA's apparent influence on cardiolipin transport seems to be a secondary consequence of its role in regulating LPS biosynthesis. In support of this idea, Clairfeuille *et al.* confirmed the finding[16] that PbgA was required for the outer membrane to retain its integrity, whereas eliminating cardiolipin had no effect.

Clairfeuille and colleagues' key advance was to analyse the structure of PbgA at a resolution of 1.9 ångströms, using a technique called X-ray crystallography. They found that PbgA belongs to a family of enzymes that also includes EptA — a protein that adds a phospholipid-derived molecular modification to the lipid A domain of LPS[17]. Lipid A is made of two phosphorylated sugars. By modifying these phosphate groups, EptA provides cells with resistance to antibiotics that bind to lipid A, called polymyxins.

The authors showed that the external surface of PbgA was tightly bound to an LPS molecule. They then re-evaluated a lower-resolution structure of PbgA[13] and — on the basis of the distance between its phosphate groups — verified that it was bound to the lipid A domain of LPS. Although a phospholipid partially occupies a site near the bound LPS, PbgA has lost the amino-acid side chains used by EptA to catalyse LPS modification. Whether or not PbgA retains enzymatic activity remains to be determined.

The picture of PbgA that emerges from Clairfeuille and colleagues' structure is of a protein that has been adapted as a receptor to sense LPS at the external surface of the inner membrane. The structure supports the model that a PbgA–LapB–FtsH–LpxC regulatory circuit acts as a control mechanism, modulating LPS biosynthesis to meet the physical demands of the cell's interconnected double membranes. Indeed, the researchers also confirm the finding[4] that a direct physical interaction occurs between PbgA and LapB in membranes. But how LPS–PbgA binding relaxes the inhibition that PbgA exerts on the LapB–FtsH interaction remains unknown.

Clairfeuille and co-workers' structure reveals that PbgA binds the lipid A moiety through a linker domain, using an amino-acid sequence that has not been reported in any other LPS-binding protein. Mutations in this LPS-binding motif compromised PbgA function. In a final set of experiments, the authors demonstrated that a synthetic peptide based on this sequence could bind LPS and inhibit bacterial growth. Through rational design, they improved the peptide's antibiotic spectrum and potency.

The polymyxins bind lipid A by interacting with both of its phosphorylated sugars[18], but PbgA binds to just one. The polymyxin antibiotic colistin is used as a last resort for treatment of infections in the clinic, but it can also increase outer membrane permeability, thereby sensitizing bacteria to more-effective antibiotics[18]. Clairfeuille and co-workers' show that the PbgA-derived peptide also sensitizes bacteria to other antibiotics, acts in synergy with colistin, and is not hampered by the LPS modifications catalysed by EptA.

PbgA was one of the few essential proteins in *E. coli* without a well-characterized function[4]. The discovery that PbgA is the LPS signal transducer provides insights for antibiotic development, in addition to illuminating a remarkable lipid balancing act in the bacterial membrane.

Russell E. Bishop is in the Department of Biochemistry and Biomedical Sciences, and at the Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario L8S 4K1, Canada.
e-mail: bishopr@mcmaster.ca

1. Bertani, B. & Ruiz, N. *EcoSal Plus* https://doi.org/10.1128/ecosalplus.ESP-0001-2018 (2018).
2. Clairfeuille, T. *et al. Nature* **584**, 479–483 (2020).
3. Guest, R. L., Guerra, D. S., Wissler, M., Grimm, J. & Silhavy, T. J. *mBio* **11**, e00598-20 (2020).
4. Fivenson, E. M. & Bernhardt, T. G. *mBio* **11**, e00939-20 (2020).
5. Nguyen, D., Kelly, K., Qiu, N. & Misra, R. *J. Bacteriol.* https://doi.org/10.1128/JB.00303-20 (2020).
6. Xie, R., Taylor, R. J. & Kahne, D. *J. Am. Chem. Soc.* **140**, 12691–12694 (2018).
7. Nikaido, H. *Chem. Biol.* **12**, 507–509 (2005).
8. Jia, W. *et al. J. Biol. Chem.* **279**, 44966–44975 (2004).
9. Wu, T. *et al. Proc. Natl Acad. Sci. USA* **103**, 11754–11759 (2006).
10. Malinverni, J. C. & Silhavy, T. J. *Proc. Natl Acad. Sci. USA* **106**, 8009–8014 (2009).
11. Chong, Z.-S., Foo, W.-F. & Chng, S.-S. *Mol. Microbiol.* **98**, 1133–1146 (2015).
12. Dalebroux, Z. D. *et al. Cell Host Microbe* **17**, 441–451 (2015).
13. Fan, J., Petersen, E. M., Hinds, T. R., Zheng, N. & Miller, S. I. *mBio* **11**, e03277-19 (2020).
14. Helander, I. M., Hirvas, L. Tuominen, J. & Vaara, M. *Eur. J. Biochem.* **204**, 1101–1106 (1992).
15. Helander, I. M., Lindner, B., Seydel, U. & Vaara, M. *Eur. J. Biochem.* **212**, 363–369 (1993).
16. Qiu, N. & Misra, R. *J. Bacteriol.* **201**, e00340-19 (2019).
17. Anandan, A. *et al. Proc. Natl Acad. Sci. USA* **114**, 2218–2223 (2017).
18. Vaara, M. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2019.01689 (2019).

Electronics

One-way supercurrent achieved in a polar film

Toshiya Ideue & Yoshihiro Iwasa

Diodes are devices that conduct electric current mainly in one direction. An electrically polar film that acts as a diode for superconducting current could lead to electronic devices that have ultralow power consumption. **See p.373**

An essential process in modern electronics is rectification, whereby bidirectional electric current is converted to unidirectional current. Electronic devices that enable rectification are called diodes and are widely used to transform alternating current into direct current, protect electric circuits from excess voltage and detect electromagnetic waves. Extending this concept to a superconducting current, which flows with zero resistance, is a fascinating challenge from both fundamental and technological viewpoints. On page 373, Ando *et al.*[1] report the achievement of this superconducting diode effect and its magnetic control in an electrically polar film that is non-centrosymmetric — lacking symmetry under a transformation known as spatial inversion. The authors' findings demonstrate that charge can be transported in a single direction without energy loss.

In a conventional diode, rectification is realized using a heterojunction (an interface between two different semiconductors), such as a p–n junction (Fig. 1a). For a p–n junction, one of the semiconductors is p-type, containing an excess of positively charged electron vacancies called holes, and the other is n-type, containing an excess of negatively charged electrons. Electric current flows easily only from one side of the interface to the other[2]. Although such a structure is a fundamental component of many devices today, it is difficult to achieve the superconducting-diode effect by this strategy because a non-zero electrical resistance at the junction is inevitable.

Non-centrosymmetric conductors can exhibit an intrinsic rectification effect, even if they are uniform and junction-free

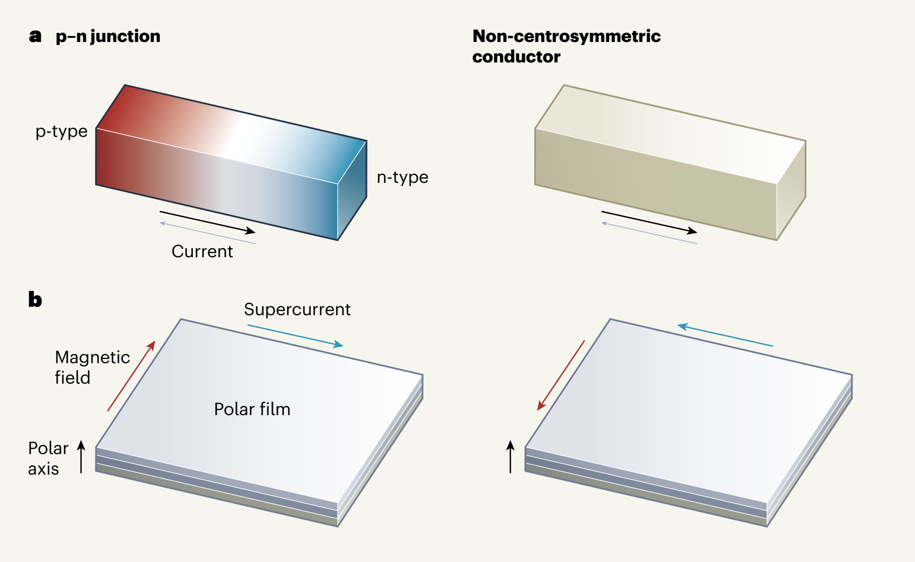


Figure 1 | Different types of rectification. a, Rectification is a process that causes electric current to flow freely in one direction but only slightly (or not at all) in the opposite direction. This process can be realized at a p–n junction, which is the interface between two types of semiconductor known as p-type and n-type. It can also be achieved in an electrical conductor that is junction-free and non-centrosymmetric – lacking symmetry under a transformation known as spatial inversion. **b**, Ando *et al.*[1] made an electrically polar film that consists of stacked layers of three metals. The authors applied a magnetic field perpendicular to the polar axis of the film and observed a superconducting current in a single direction perpendicular to the directions of both the magnetic field and the polar axis. They found that the direction of this rectified supercurrent could be inverted by reversing the direction of the magnetic field.

(Fig. 1a). This effect is currently recognized as a fundamental feature of these materials and as an emergent physical property that reflects the characteristic electronic states, magnetic structure, interaction effects and geometric or topological nature of electrons in non-centrosymmetric solids.

If this intrinsic rectification effect occurs alongside broken time-reversal symmetry (a lack of symmetry when the direction of time is reversed), it is known as magnetochiral anisotropy. Since this phenomenon was first reported[3] in 2001, it has been studied in a variety of quantum materials and interface systems[4–7]. A key aspect of magnetochiral anisotropy is that, in principle, it can occur in any quantum phase of matter, including a superconducting phase under appropriate symmetry conditions. Moreover, the direction of the rectified current can be inverted by reversing the direction of the magnetic field or magnetization.

In 2017, scientists observed magnetochiral anisotropy in two-dimensional non-centrosymmetric superconductors[4]. They suggested that the effect is a hallmark of exotic superconducting states, such as those in which the Cooper pairs (the electron pairs responsible for superconductivity) have an unconventional pairing symmetry. Therefore, magnetochiral anisotropy could provide a powerful experimental probe of non-centrosymmetric superconductors[4]. Moreover, a relatively large rectification effect has been detected in superconducting films that have microstructures, such as triangular magnets through the motion of vortices[5] – magnetic fluxes that pierce superconductors. However, the realization of an ideal superconducting rectifier, in which the zero-resistance state is retained in only one direction, has been both lacking and highly anticipated.

Ando *et al.* produced an artificial film called a superlattice that is composed of stacked alternating layers of niobium, vanadium and tantalum. The superlattice has an electrically polar structure because mirror symmetry along the stacking direction is broken. The authors focused on electric transport along the film's plane, which is uniform and junction-free. In previous studies on interfaces[6] and polar crystals[7], an intrinsic rectification effect was observed along the plane when a magnetic field was applied perpendicular to both the current and the polar axis. Using a similar set-up, Ando and colleagues detected ideal superconducting diode behaviour in their film (Fig. 1b).

> **"The authors' work opens the door to a new era of superconductivity research."**

Because the authors' film is relatively thick (120 nanometres), it can be regarded as a 3D superconductor. It shows a sharp transition between conducting and superconducting states when it is cooled to temperatures below 4.4 kelvin, which is needed for the current to completely switch between these states. Moreover, the direction of the rectified current can be reversed by inverting the direction of the magnetic field, which is useful for practical applications (Fig. 1b).

The authors' results indicate the great potential of non-centrosymmetric superconductors for producing devices that have ultrahigh sensitivity to electromagnetic fields or ultralow power consumption. The findings could also pave the way to unexpected device capabilities that are even more intriguing. The use of a superlattice is advantageous because the superconducting-diode effect should be controllable by tuning the superlattice's structure. For example, by choosing appropriate constituent elements and optimizing the film's thickness or number of stacked layers, it might be possible to obtain samples that have, relative to the authors' film, a higher superconducting transition temperature or a higher resistance in the opposite direction to that of the rectified current; such samples would be desirable for applications. Another possibility is that the direction of the rectified current could be reversed by merely inverting the stacking order.

An important future issue is to clarify and fully understand the superconducting state in this superlattice and the microscopic mechanism of the superconducting diode effect. Ando *et al.* focus on a well-documented interaction in polar systems, known as the Rashba effect, and discuss the possible impact of the unconventional pairing symmetry in the superconducting state. However, there might be other contributions to the film's behaviour from vortex motion or electron-scattering processes. Despite these remaining issues, there is no doubt that the authors' work opens the door to a new era of superconductivity research.

Toshiya Ideue and **Yoshihiro Iwasa** are at the Quantum-Phase Electronics Center and in the Department of Applied Physics, University of Tokyo, Tokyo 113-8656, Japan.
e-mails: ideue@ap.t.u-tokyo.ac.jp; iwasa@ap.t.u-tokyo.ac.jp

1. Ando, F. *et al. Nature* **584**, 373–376 (2020).
2. Sze, S. M. *Semiconductor Devices: Physics and Technology* (Wiley, 1981).
3. Rikken, G. L. J. A., Fölling, J. & Wyder, P. *Phys. Rev. Lett.* **87**, 236602 (2001).
4. Wakatsuki, R. *et al. Sci. Adv.* **3**, e1602390 (2017).
5. Villegas, J. E. *et al. Science* **302**, 1188–1191 (2003).
6. Rikken, G. L. J. A. & Wyder, P. *Phys. Rev. Lett.* **94**, 016601 (2005).
7. Ideue, T. *et al. Nature Phys.* **13**, 578–583 (2017).

Genomics

The remarkable tuatara finds its place

Rebecca N. Johnson

The genome sequence of an unusual reptile called the tuatara sheds light on the species' evolution and on conservation strategies. The work is a model of current best practice for collaborating with Indigenous communities. **See p.403**

A once-species-rich order of reptiles called the Rhynchocephalia lived across the globe during the time of the dinosaurs[1,2]. Just one of these species survives today: the tuatara (Fig. 1). Found only in New Zealand, tuatara are a *taonga* ('special treasure') for Maori people. The reptiles have a set of intriguing traits – including longevity and an unusual combination of bird- and reptile-like morphological features[3] – that have led to uncertainty over their place in the evolutionary tree. On page 403, Gemmell *et al.*[4] report the first whole-genome sequence for the tuatara (*Sphenodon punctatus*). The researchers' study provides insights into the biology and evolution of this extraordinary animal.

The work is a collaboration between genomicists and Ngātiwai, the Maori *iwi* (people) who have guardianship over the tuatara populations used in this study. Even with the advances in genome-sequencing technology over the past several years, it is not possible to produce a high-quality genome sequence without access to good genetic material. The researchers obtained this only through collaboration. Ngātiwai were involved in all decision-making processes for this study, and are commendably listed as the paper's last authors. Gemmell *et al.* also provide a template agreement that other researchers can follow should they wish to consult with traditional guardians of other organisms. As such, the study sets a new standard for collaboration with Indigenous guardians on genomics and other scientific endeavours.

The genome produced by Gemmell and co-workers is one of the largest vertebrate genomes published so far. At more than 5 gigabases, it is about 50% larger than the human genome. To complement the genome, the authors generated gene-expression profiles for tuatara blood and embryos. They also performed a preliminary analysis of active and inactive sections of the genome, and an in-depth analysis of repeated regions. The genome represents a valuable resource for future research into a variety of topics – from the evolution of egg laying to why the once-species-rich Rhynchocephalia has only a single survivor.

One reason for sequencing genomes is to reconstruct the evolutionary tree of life; this allows a deeper understanding of how life evolved, and this knowledge can be used to tackle challenges such as biodiversity loss and climate change. Gemmell *et al.* used comparative-genomics methods to do just that.

PETE OXFORD/NPL

Figure 1 | A tuatara in New Zealand. Gemmell *et al.*[4] have generated a high-quality genome sequence for the tuatara (*Sphenodon punctatus*).

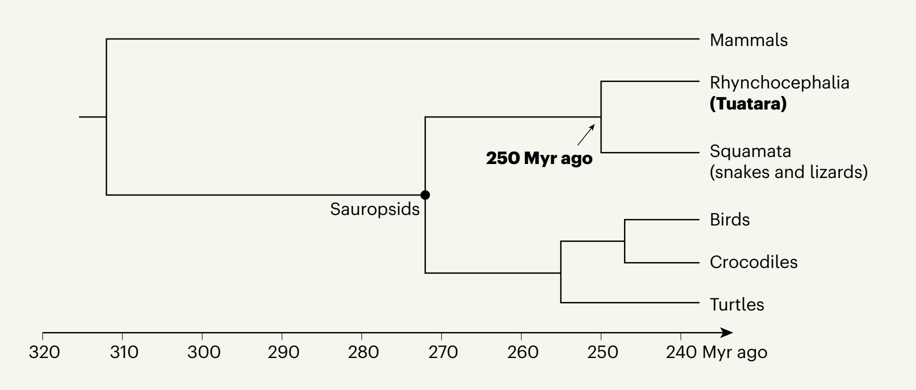


Figure 2 | Refining the evolutionary tree for reptiles, birds and mammals. This phylogenetic tree includes six branches: mammals and five branches within a clade called sauropsids, which comprises reptiles and birds. One of these, the Rhynchocephalia, has only one living member, the tuatara. Gemmell and colleagues date the divergence of the Rhynchocephalia from the Squamata to about 250 million years (Myr) ago.

They generated a phylogenetic tree for the Sauropsida (a clade that includes all modern reptiles, along with birds) by comparing the genome sequences of 27 vertebrates, including the tuatara (Fig. 2). The researchers' tree confirms a previous suggestion[5] that the Rhynchocephalia diverged from their closest relatives[5,6], the Squamata (lizards and snakes), about 250 million years ago, during the Early Triassic period. Confirmation of such an early divergence is important for understanding the origin and evolution of the Lepidosauria, which comprises both the Rhynchocephalia and the Squamata.

Could the tuatara be a living fossil? The term, which refers to a species that has evolved extremely slowly and still retains the features of its ancient ancestors, has fallen out of favour with palaeontologists and evolutionary biologists. This is due, in part, to misuse of the term, which can arise when fossil evidence that would have reflected physical changes in a species is missing, or when researchers mistakenly assume that a lone survivor of a given lineage must have remained unchanged over evolutionary time. Tuatara have a close resemblance to their forebears from the early Mesozoic era[7], between 240 million and 230 million years ago. However, there is no continuous fossil record[6], making it difficult to define which traits the tuatara might share with its now-extinct ancestors.

Gemmell and colleagues' phylogenetic reconstruction indicates that the tuatara has the lowest rate of evolution of any lepidosaur described so far. These data could suggest that the tuatara is indeed a living fossil. In addition to its long generation time and low body temperature, the tuatara's slow evolution could make it particularly vulnerable to a warming climate.

The authors then analysed the tuatara's genome in more detail. On average, more than 50% of a vertebrate's genome is comprised of repetitive DNA sequences (repeat elements)[8,9]. In line with this figure, 64% of the tuatara genome is repeat elements. However, the types of repeat element were a combination of mammal-like and reptile-like. This is a key finding, because the most-recent common ancestor of sauropsids was imputed to be reptile-like on the basis of genomic features found in birds and lizards, some of which have very well-characterized genomes[10]. By revealing unexpected, mammal-like features, the tuatara genome provides new evolutionary insights.

The researchers also found that the tuatara genome has a broader range of DNA sequences called transposons (sequences that can move from one genomic location to another) than has any other reptile, bird or mammal sequenced so far. Many of these seem to have been active recently (probably in the past few million years), suggesting that they still have or have recently had a role in shaping the genome. The authors suggest that the tuatara's large genome might be explained by the fact that almost one-third of it consists of duplications of DNA sequences between 1 and 400 kilobases long.

> **"The study sets a new standard for collaboration with Indigenous guardians on genomics and other scientific endeavours."**

Gemmel *et al.* then compared tuatara genes associated with eyesight, smell, immunity, thermoregulation and longevity with the equivalent genes in other species. Despite being nocturnal, the tuatara is a highly visual predator, and the authors found evidence that it has retained vision-associated genes reminiscent of an ancestor that was active during the day. The species seems to have retained robust colour vision, even at low light levels – suggesting that there could be an adaptive benefit to having this trait.

In addition, tuatara seem to have a repertoire of several hundred odour receptors – similar to the number in birds, but lower than that in crocodiles or turtles. Further research is required to investigate the function of these receptors and to determine the implications of this reduced receptor repertoire for tuatara feeding and hunting. For instance, perhaps tuatara rely on their vision for hunting (like birds), rather than depending on odours and other senses (as do snakes).

Finally, there is an ongoing debate about whether there are actually two subspecies of tuatara – crucial information for conservation strategies. Because the animals are protected, the authors assessed genetic diversity among the population using samples collected over many decades. This analysis confirms that there is only one species of tuatara, despite one population (on North Brother Island in the Cook Strait) being genetically distinct from the others. The lack of current samples is not desirable for designing genetics-based conservation approaches, but, given the tuatara's longevity, any recommendations arising from the study are still likely to be valid.

Much as whole-genome sequencing has benefited human health and improved our understanding of human evolution, the sequencing of genomes of other organisms can have many benefits – especially for those organisms facing biodiversity loss caused by humans. However, for many such species, samples are not readily available. Gemmell and colleagues' work reminds us that sample collection and consultation with Indigenous people can go hand in hand to improve outcomes for both biological and cultural conservation.

Rebecca N. Johnson is at the Smithsonian Institution, National Museum of Natural History, Washington, DC 20560, USA.
e-mail: johnsonrn@si.edu

1. Evans, S. E. *Biol. Rev. Camb. Phil. Soc.* **78**, 513–551 (2003).
2. Hsiou, A. S. *et al. Sci. Rep.* **9**, 11821 (2019).
3. Gunther, A. *Phil. Trans. R. Soc. Lond.* **157**, 595–629 (1867).
4. Gemmell, N. J. *et al. Nature* **584**, 403–409 (2020).
5. Jones, M. E. H. *et al. BMC Evol. Biol.* **13**, 208 (2013).
6. Jones, M. E. H. & Cree, A. *Curr. Biol.* **22**, R986–R987 (2012).
7. Herrera-Flores, J. A., Stubbs, T. L. & Benton, M. J. *Palaeontology* **60**, 319–328 (2017).
8. Pasquesi, G. I. M. *et al. Nature Commun.* **9**, 2774 (2018).
9. Sotero-Caio, C. G., Platt, R. N. II, Suh, A. & Ray, D. A. *Genome Biol. Evol.* **9**, 161–177 (2017).
10. Zhang, G. *et al. Science* **346**, 1311–1320 (2014).

Perspective

A perspective on potential antibody-dependent enhancement of SARS-CoV-2

https://doi.org/10.1038/s41586-020-2538-8

Received: 15 May 2020

Accepted: 6 July 2020

Ann M. Arvin[1,2]✉, Katja Fink[1,3], Michael A. Schmid[1,3], Andrea Cathcart[1], Roberto Spreafico[1], Colin Havenar-Daughton[1], Antonio Lanzavecchia[1,3], Davide Corti[1,3] & Herbert W. Virgin[1,4]✉

Antibody-dependent enhancement (ADE) of disease is a general concern for the development of vaccines and antibody therapies because the mechanisms that underlie antibody protection against any virus have a theoretical potential to amplify the infection or trigger harmful immunopathology. This possibility requires careful consideration at this critical point in the pandemic of coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Here we review observations relevant to the risks of ADE of disease, and their potential implications for SARS-CoV-2 infection. At present, there are no known clinical findings, immunological assays or biomarkers that can differentiate any severe viral infection from immune-enhanced disease, whether by measuring antibodies, T cells or intrinsic host responses. In vitro systems and animal models do not predict the risk of ADE of disease, in part because protective and potentially detrimental antibody-mediated mechanisms are the same and designing animal models depends on understanding how antiviral host responses may become harmful in humans. The implications of our lack of knowledge are twofold. First, comprehensive studies are urgently needed to define clinical correlates of protective immunity against SARS-CoV-2. Second, because ADE of disease cannot be reliably predicted after either vaccination or treatment with antibodies—regardless of what virus is the causative agent—it will be essential to depend on careful analysis of safety in humans as immune interventions for COVID-19 move forward.

The benefit of passive antibodies in ameliorating infectious diseases was recognized during the 1918 influenza pandemic[1]. Since then, hyperimmune globulin has been widely used as pre- and post-exposure prophylaxis for hepatitis A, hepatitis B, chickenpox, rabies and other indications for decades without evidence of ADE of disease[2] (see Box 1 for definition of terms). The detection of antibodies has also been a reliable marker of the effectiveness of the many licensed human vaccines[3]. The antiviral activity of antibodies is now known to be mediated by the inhibition of entry of infectious viral particles into host cells (neutralization) and by the effector functions of antibodies as they recruit other components of the immune response. Neutralizing antibodies are directed against viral entry proteins that bind to cell surface receptors, either by targeting viral proteins that are required for fusion or by inhibiting fusion after attachment[4–6] (Fig. 1). Antibodies can cross-neutralize related viruses when the entry proteins of the viruses share epitopes—the part of a protein to which the antibody attaches. Antibodies also eliminate viruses through effector functions triggered by simultaneous binding of the antigen-binding fragment (Fab) regions of immunoglobulin G (IgG) to viral proteins on the surfaces of viruses or infected cells, and of the fragment crystallizable (Fc) portion of the antibody to Fc gamma receptors (FcγRs) that are expressed by immune cells[7,8] (Fig. 2). Antibodies that mediate FcγR- and complement-dependent effector functions may or may not have neutralizing activity, can recognize other viral proteins that are not involved in host-cell entry and can be protective in vivo independent of any Fab-mediated viral inhibition[9,10]. Recent advances in FcR biology have identified four activating FcγRs (FcγRI, FcγRIIa, FcγRIIc and FcγRIIIa) and one inhibitory FcγR (FcγRIIb) that have various Fc ligand specificities and cell-signalling motifs[10]. The neonatal Fc receptor (FcRn) has been described to support antibody recycling and B and T cell immunity through dendritic cell endocytosis of immune complexes[11,12]. Natural killer cells recognize IgG–viral protein complexes on infected cells via FcγRs to mediate antibody-dependent cytotoxicity, and myeloid cells use these interactions to clear opsonized virions and virus-infected cells by antibody-dependent cellular phagocytosis (Fig. 2). The complement pathway is also activated by Fc binding to the complement component C1q, resulting in the opsonization of viruses or infected cells and the recruitment of myeloid cells. Antibody effector functions also contribute to antiviral T-cell-mediated immunity in vivo[13]. Notably, new knowledge about Fc effector functions has led to improved passive-antibody therapies through Fc modifications that reduce or enhance interactions with FcγRs, lengthen the half-life of the antibody and potentially enhance antigen presentation to T cells, providing what is termed a vaccinal effect[8,11,14].

Although their importance for protection is indisputable, the concern about ADE of disease arises from the possibility that antibodies

---

[1]Vir Biotechnology, San Francisco, CA, USA. [2]Stanford University School of Medicine, Stanford, CA, USA. [3]Humabs Biomed SA, a subsidiary of Vir Biotechnology, Bellinzona, Switzerland. [4]Washington University School of Medicine, Saint Louis, MO, USA. ✉e-mail: aarvin@vir.bio; svirgin@vir.bio

Box 1

Definitions

ADE of disease: Enhancement of disease severity in an infected person or animal when an antibody against a pathogen—whether acquired by an earlier infection, vaccination or passive transfer—worsens its virulence by a mechanism that is shown to be antibody-dependent.

Vaccine enhancement of disease: Enhancement of disease severity in an infected person or animal that had been vaccinated against the pathogen compared to unvaccinated controls. This results from deleterious T cell responses or ADE of disease and is usually difficult to link to one or the other.

Neither ADE of disease nor vaccine enhancement of disease have established, objective clinical signs or biomarkers that can be used to distinguish these events from severe disease caused by the pathogen. Carefully controlled human studies of sufficient size enable the detection of an increased frequency of severe cases in cohorts given passive antibodies or vaccines compared to the control group, and atypical manifestations of infection can be identified should they occur.

Mechanisms of antibody-mediated protection and the potential for ADE of infection

The essential benefits of antibodies are mediated by several well-defined mechanisms that also have the potential for ADE of infection. Protection as well as ADE of infection can be observed in various assays of virus–cell interactions. An observation of ADE of infection in vitro does not predict ADE of disease in humans or animals.

Virus entry: Antibodies block viruses by interfering with their binding to receptors on host cells or inhibiting changes in the viral protein needed for entry.

Virus binding and internalization: Antibodies bind viruses to cells of the immune system via Fcγ receptors on the cell surface and internalization of viruses typically results in their degradation.

Instead of protection, ADE of infection may occur if antibody binding improves the capacity of the viral protein to enable entry of the virus into its target cell, or if the virus has the capacity to evade destruction and produce more viruses after Fcγ receptor-mediated entry.

Cytokine release: Antibodies that bind viruses and Fcγ receptors on cells of the immune system trigger the release of cytokines that inhibit viral spread and recruit other immune cells to eliminate infected cells. Although a part of the normal protective immune response, this can result in ADE of disease if excessive.

Complement activation: Antibodies binding to virus or viral proteins on host cells may activate the complement cascade, a series of plasma proteins that together have a role in protective immunity through multiple mechanisms. Formation of large complexes of antibodies and viral proteins (antigens) can lead to immune complex deposition that activates complement. When excessive, antibody-dependent activation of complement may result in tissue damage and potential ADE of disease.

Antibody-mediated mechanisms in the development of memory immunity: Antibodies bound to viruses or viral proteins can be taken up Fcγ receptors into immune system cells that process the antigens for activation and expansion of B cells and T cells. These mechanisms, which are critical for the establishment of memory immunity against future encounters with the virus, balance the potential risk of amplification of infection after viral uptake by some immune system cells.

present at the time of infection may increase the severity of an illness. The enhancement of disease by antibody-dependent mechanisms has been described clinically in children given formalin-inactivated respiratory syncytial virus (RSV) or measles vaccines in the 1960s, and in dengue haemorrhagic fever due to secondary infection with a heterologous dengue serotype[15–21]. For example, antibodies may enable viral entry into FcγR-bearing cells, bypassing specific receptor-mediated entry; this is typically followed by degradation of the virus, but could amplify infection if progeny virions can be produced. Although cytokine release triggered by interactions between the virus, antibody and FcγR is also highly beneficial—owing to direct antiviral effects and the recruitment of immune cells—tissue damage initiated by viral infection may be exacerbated[22].

While recognizing that other mechanisms of immune enhancement may occur, the purpose of this Perspective is to review clinical experiences, in vitro analyses and animal models relevant to understanding the potential risks of antibody-dependent mechanisms and their implications for the development of the vaccines and antibodies that will be essential to stop the COVID-19 pandemic. Our objective is to evaluate the hypothesis that antibody-mediated enhancement is a consequence of low-affinity antibodies that bind to viral entry proteins but have limited or no neutralizing activity; antibodies that were elicited by infection with or vaccination against a closely related serotype, termed 'cross-reactive' antibodies; or suboptimal titres of otherwise potently neutralizing antibodies. We assess whether there are experimental approaches that are capable of reliably predicting ADE of disease in humans and conclude that this is not the case.

Principles for assessing potential ADE of disease

The use of ADE to denote enhanced severity of disease must be rigorously differentiated from ADE of infection—that is, from the binding, uptake and replication of the virus, cytokine release or other activities of antibodies detected in vitro. The first principle is that an antibody-dependent effect in vitro does not represent or predict ADE of disease without proof of a role for the antibody in the pathogenesis of a more severe clinical outcome. A second principle is that animal models for the evaluation of human polyclonal antibodies or monoclonal antibodies (mAbs) should be judged with caution because FcRs that are engaged by IgGs are species-specific[23,24], as is complement activation. Antibodies can have very different properties in animals that are not predictive of those in the human host, because the effector functions of antibodies are altered by species-specific interactions between the antibody and immune cells. Animals may also develop antibodies against a therapeutic antibody that limit its effectiveness, or cause immunopathology. In addition, the pathogenesis of a model virus strain in animals does not fully reflect human infection because most viruses are highly species-specific. These differences may falsely support either protective or immunopathological effects of vaccines and antibodies. A third principle is that the nature of the antibody response depends on the form of the viral protein that is recognized by the immune system, thus determining what epitopes are presented. Protective and non-protective antibodies can be elicited to different forms of the same protein. A fourth principle is that mechanisms of pathogenesis in the human host differ substantially among viruses, or even between strains of a particular virus. Therefore, findings regarding the effects of passive antibodies or vaccine-induced immunity on outcomes cannot be extrapolated with confidence from one viral pathogen to another.

Observations about RSV, influenza and dengue

As background for considering the risks of ADE of disease caused by SARS-CoV-2, it is important to closely examine clinical circumstances relevant to the hypothesis that antibodies predispose to ADE of disease

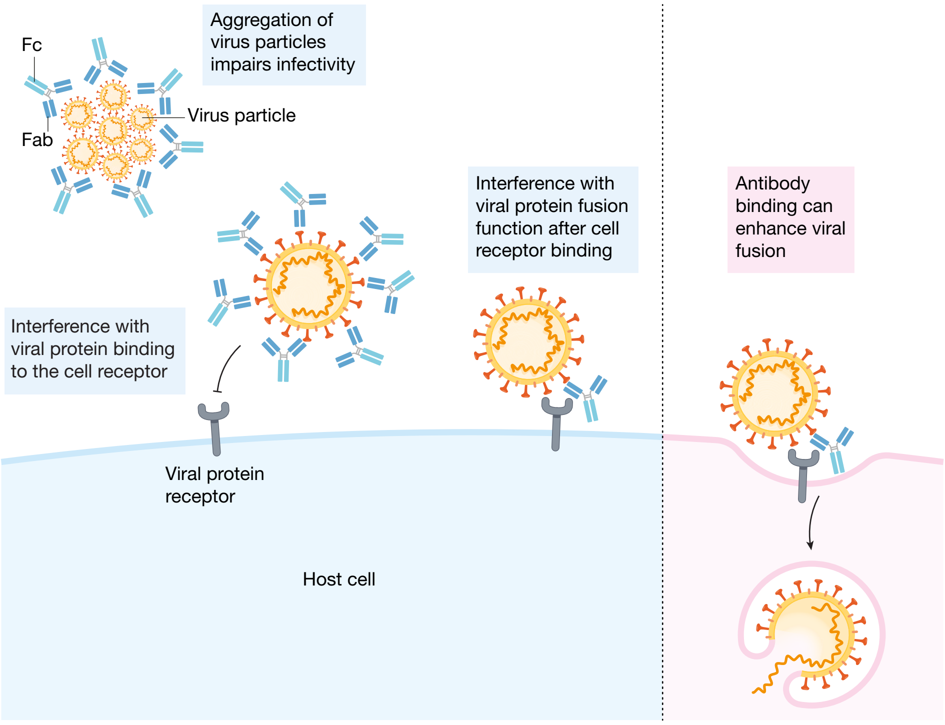


Fig. 1 | Neutralization of viruses by functions of the IgG Fab fragment. Mechanisms of antibody-mediated neutralization of viruses by functions of the IgG Fab fragment that block binding to cell surface receptors and inhibit infectivity by aggregating viral particles and inhibiting steps in the viral life cycle, such as fusion. Binding of antibodies with certain properties may enable changes in the viral entry protein that accelerate fusion.

by amplifying infection or through damaging inflammatory responses. We focus on the clinical experiences with RSV, influenza and dengue to demonstrate the complexities of predicting from in vitro assays or animal models whether passively transferred or vaccine-induced antibodies will cause ADE of disease, and of differentiating ADE from a severe illness that is unrelated to pre-existing antibodies.

RSV

In a study of RSV in children under the age of 2 years, there were more cases requiring hospitalization for RSV-related bronchiolitis or pneumonia—especially in those aged between 6 and 11 months—in children who were immunized with a formalin-inactivated (FI)-RSV vaccine (10/101) than in children who were not immunized with FI-RSV (control cases; 2/173)[25]. This was also observed in a second study (18/23 hospitalizations of immunized children, with two deaths, compared with 1/21 control cases)[16] and in two smaller studies[17,26]. This condition has been termed vaccine-associated enhanced respiratory disease. Later studies showed that the ratio of fusion protein (F) binding antibodies to neutralizing antibodies was higher in the sera of 36 vaccinated compared to 24 naturally infected children, suggesting that non-neutralizing antibodies to an abnormal F-protein conformation may have been a predisposing factor[27]. Complement activation, detected by the presence of C4d in the lungs of the two fatal cases, suggested that antibody–F protein immune complexes led to more severe disease[28]. However, C4d deposition can result from the lectin-binding pathway as well as from the classical pathway, and C4 can be produced by epithelial cells and activated by tissue proteases[29]. Whether harmful RSV-specific T cells were induced was not determined: although lymphocyte transformation frequencies were higher, this early method did not differentiate antigen-specific responses from secondary cytokine stimulation or from CD4 and CD8 T cell responses, although CD4 T cell proliferation is more likely[30]. Importantly, the FI-RSV clinical experience did not establish that vaccine-enhanced disease was antibody-dependent[31]. Subsequently, in animal studies, the production of low-avidity antibodies due to insufficient Toll-like-receptor signalling and lack of antibody maturation, and the formation of immune complexes have been implicated. However, a definitive antibody-mediated mechanism of enhancement has not been documented[32], and models have also identified Th2-skewing of the T cell response and lung eosinophilia with challenge after FI-RSV, raising the possibility that T cells contribute to vaccine-induced enhancement of RSV disease[31,33].

Experience with RSV also includes more than 20 years of successful prophylaxis of high-risk infants with palivizumab, a mAb directed against pre- and post-fusion F protein[34]. Importantly, this experience challenges a role for low neutralizing-antibody titres in the ADE of lung disease, because RSV morbidity does not increase as titres decrease. Further, if suboptimal neutralization were a factor, the failure of suptavumab—caused by F protein drift in RSV B strains—would be associated with ADE of disease; however, infections in such cases were not more severe[35]. Clinical trials of an RSV mAb that has an extended half-life have shown a reduction in hospitalizations of around 80%, again supporting the concept that such treatments provide protection without a secondary risk from declining titres[36]. mAbs against RSV have been consistently safe, even as the neutralizing capacity diminishes after administration.

Influenza

Influenza is instructive when considering the hypothesis that cross-reactive antibodies predispose to ADE of disease, because almost all humans contain antibodies that are not fully protective against antigenically drifted strains that emerge year after year. Instead, pre-existing immunity typically provides some protection against a second viral strain of the same subtype. Antibodies against neuraminidase and against the stem or head regions of haemagglutinin also correlate with protection[37]. When an H1N1 strain with a haemagglutinin shift emerged in the 2009 H1N1 pandemic, some epidemiological studies linked a greater incidence of medically treated illness to previous vaccination against influenza, whereas others did not[38–41]. One report correlated cross-reactive, low-avidity and poorly neutralizing antibodies with risk in middle-aged people—the demographic with a higher prevalence of severe 2009 H1N1[42]. Immunopathology and C4d were reported in the lungs of six fatal cases in this age group, indicating that antibody-dependent complement activation through immune-complex formation may have been a contributing factor. However, as noted above, other mechanisms lead to C4d deposition, and lung T lymphocytosis attributed to T cell epitopes shared by 2009 H1N1 and earlier H1N1 strains was also observed, raising the possibility that T cells played a part. Another study correlated pre-existing antibodies that mediated infected cell lysis by complement activation with protection against H1N1 in children[43]. In a porcine model, enhanced pulmonary disease was observed after vaccination with an inactivated influenza H1N2 strain followed by heterologous H1N1 challenge[44]. The animals had non-neutralizing antibodies that bound haemagglutinin in the stem region, but did not block the binding of haemagglutinin to its cell receptor and accelerated fusion in vitro by a Fab-dependent mechanism (Fig. 1). Lung pathology was also observed in mice treated with a mAb that induced a conformational change in haemagglutinin that facilitated fusion[45]. Such a mechanism was postulated to have potential clinical relevance when the infecting influenza virus has undergone antigenic shift and the infection boosts non-neutralizing haemagglutinin-stem-binding antibodies without a neutralizing antibody response. The likelihood of these circumstances occurring is unclear. Further, human influenza vaccines are not known to elicit immunodominant antibodies with this property. Importantly, as noted above, stem antibodies correlate both with resistance to infection and to severe disease in humans, indicating that this interesting mechanism is not predictive of disease causation for stem-specific antibodies[37]. In addition, mAbs can be screened to avoid fusion-enhancing properties, and fusion is not intrinsically accelerated by low titres of neutralizing antibodies. Notably, infants benefit from immunization from six months of age, despite their limited capacity to produce affinity-matured, high-avidity antibodies. Overall, widespread annual surveillance of influenza does not reveal ADE of disease, even though cross-reactive strains and vaccine mismatches are common.

Dengue

There are four viral serotypes of dengue that circulate in endemic areas[19]. Although severe dengue haemorrhagic fever and shock syndrome occurs during primary infection, possible ADE of disease has been associated with poorly neutralizing cross-reactive antibodies against a heterologous dengue serotype. Taking into account the difficulty of classification due to the overlapping signs of severe infection and ADE of disease, clinical experience indicates that ADE of disease does occur, but is rare in endemic areas (36/6,684 participants; around 0.5%) and is correlated with a narrow range of low pre-existing antibody titres (1:21–1:80)[20]. In the same study, high antibody titres were found to be protective. The challenge of predicting how to avoid such a rare immune-enhancing situation against the background of protection conferred by dengue neutralizing antibodies implies that it will be equally difficult for SARS-CoV-2.

When considering conditions that may result in ADE of disease, it is important to emphasize that dengue differs from other viruses because it targets monocytes, macrophages and dendritic cells and can produce progeny virus in these cells, which abundantly express both viral entry receptors and FcγRs. ADE of infection can be demonstrated in vitro with FcγR-expressing cells—typically with cross-reactive antibodies that have low or no neutralizing activity, have low affinity, or target non-protective epitopes, or if a narrow range of antibody and infectious virus concentrations is tested[46,47]. The mechanism of ADE of disease associated with dengue therefore depends on three factors: the circulation of multiple strains of a virus that have variable antigenicity, a virus that is capable of replication in FcγR-expressing myeloid cells and sequential infection of the same person with these different viral serotypes. Despite these pre-disposing conditions and the fact that dengue is an increasingly common infectious disease, severe dengue disease is rare.

The role of pre-existing immunity has also been a concern for the quadrivalent live attenuated dengue vaccine (Dengvaxia), because higher hospitalization rates were observed among vaccine recipients who were initially seronegative—especially children aged between two and eight years[48]. Other explanations for this outcome include poor efficacy against serotypes 1–3, or the failure to induce cell-mediated immunity because T cells primarily recognize non-structural proteins that are not present in the chimeric vaccine. Importantly, the cause of death in 14 fatal cases of dengue could not be determined by the WHO (World Health Organization) Global Advisory Committee on Vaccine Safety, because a failure of vaccine protection could not be distinguished from immune enhancement by clinical or laboratory criteria[49]. This experience underscores how difficult it is to predict the potential for vaccine-induced antibodies or a therapeutic antibody to enhance the severity of disease, because other mechanisms of pathogenesis that result in severe disease are potentially involved—even for the well-studied case of dengue.

In other assessments of the risks and benefits of cross-reactive antibodies, infection with Zika—which, as with dengue, is a flavivirus—was less common in individuals who had previously been infected with dengue[50]. In addition, the presence of cross-reactive antibodies has been associated with improved efficacy, as measured by the responses to a yellow fever vaccine in recipients who had received a Japanese encephalitis vaccine[47], and by association of the effectiveness of Dengvaxia with seropositivity for dengue at the time of immunization[51].

In summary, these clinical experiences with RSV, influenza and dengue provide strong evidence that the circumstances that are proposed to lead to ADE of disease—including low affinity or cross-reactive antibodies with limited or no neutralizing activity or suboptimal titres—are very rarely implicated as the cause of severe viral infection in the human host. Furthermore, clinical signs, immunological assays or biomarkers that can differentiate severe viral infection from a viral infection enhanced by an immune mechanism have not been established[49,52].

Assessing the risk of ADE of disease with SARS-CoV-2

Given the complexities described above, it is sobering to take on the challenge of predicting ADE of disease caused by SARS-CoV-2. Here we consider whether clinical circumstances point to a role for antibodies with poor or no neutralizing activity in severe COVID-19, incorporating relevant experience from disease caused by the common human coronaviruses, as well as by severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome-related coronavirus (MERS-CoV).

Infection by SARS-CoV-2 is initiated by the binding of its fusion protein, the spike (S) protein, to the entry receptor, angiotensin-converting enzyme 2 (ACE2)[53–55]. Other receptors for SARS-CoV-2, such as CD147, have also been reported[56]. ACE2 is expressed on alveolar type II pneumocytes, airway epithelial cells, nasal tract goblet cells and ciliated cells, as well as on intestinal and other non-respiratory tract cells, as assessed by RNA expression[57]. On most such cells, ACE2 seems to be expressed at low levels; however, it can be upregulated by interferons[58], which could theoretically promote infection if the virus overcomes interferon-induced barriers. FcγRIIa and FcγRIIIa were detected in alveolar, bronchial and nasal-cavity epithelial cells by single-cell RNA sequencing, but both fractions of positive cells and levels of expression per cell were considerably lower than for resident myeloid and natural killer cells[59,60]. The moderate prevalence of both ACE2 and FcγRs results in poor co-occurrence, although this might be underestimated because of the dropout effect in single-cell transcriptomics. Co-expression of ACE2 and FcγRs therefore seems to be limited, which would mitigate against antibody-enhanced disease caused by SARS-CoV-2 via the dual-receptor mechanism proposed in dengue infection.

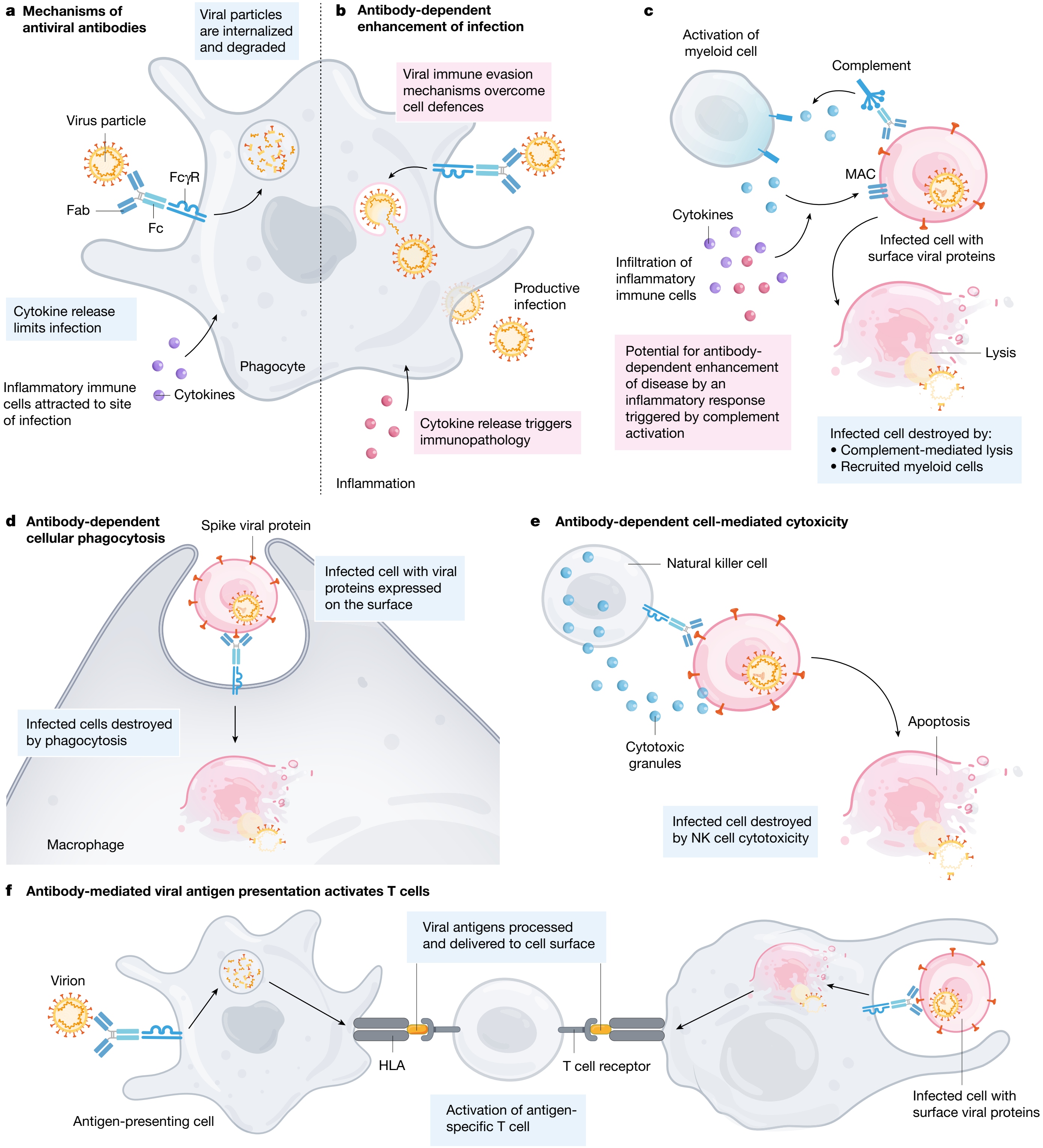


Fig. 2 | Antibody effector functions of the IgG Fc fragment. Antibody effector functions are mediated by binding of the IgG Fc domain to FcγRs on myeloid cells or to components of the complement system. These activities occur when the antibody binds the target virus protein either on virions or on infected cells. **a**, Viral particles are internalized and degraded and local cytokine release recruits immune cells. **b**, If cells are permissive, progeny virions could be produced. When virus–antibody complexes are taken up by the cell, a detrimental cytokine response may be generated. **c**, Binding of the IgG Fc fragment to C1q leads the activation of complement components C3, C3a and C5a and of the complement membrane attack complex (MAC) that disrupts membranes. C3 and C5a facilitate phagocytosis by myeloid cells. C3a and C5a are anaphylatoxins that attract inflammatory cells, which can secrete cytokines that enhance antiviral immunity but could be detrimental if produced in excess. **d**, **e**, The IgG Fc domain binds to multiple types of FcγRs on myeloid cells to trigger effector functions. The specific consequences of this interaction are dependent on the FcγR that is involved and are not detailed here. **d**, Antibody-dependent phagocytosis by macrophages and dendritic cells. **e**, Antibody-dependent cytotoxicity mediated by natural killer (NK) cells. **f**, Antibody-mediated antigen presentation after the uptake of virus or virus-infected cells by phagocytic cells leads to the activation of antiviral T cells.

When considering potential detrimental effects of antibodies, the presence or absence of cross-reactive antibodies against other human coronavirus (HCoV) strains has not been linked to whether SARS-CoV-2 infection is more severe, mild or asymptomatic, although antibodies that recognized the SARS-CoV-2 S2 subunit were detected in 12 out of 95 uninfected individuals[61]. In two reports, 30–50% of SARS-CoV-2 seronegative or unexposed individuals had CD4 T cells that recognized the SARS-CoV-2 S protein[62,63]. Previous infection with HCoV-HKU1 and HCoV-OC43 betacoronaviruses, or HCoV-NL63 and HCoV-229E alphacoronaviruses, is not known to predispose to more severe infection with the related virus from the same lineage[64–67]. Conversely, the endemic nature of coronavirus infections indicates that infection in the presence of low levels of antibodies is common, providing a theoretical opportunity for ADE of disease—although these illnesses are mild—and suggesting that cross-protection may be transient[68]. It is of interest that neither low neutralizing-antibody titres nor heterologous virus challenge were associated with enhanced disease in human studies of HCoV-229E[64,65]. Although HCoV-NL63 also uses the ACE2 entry receptor, the receptor-binding domain (RBD) of HCoV-NL63 is structurally very different from that of SARS-CoV-2, which would limit antibody cross-reactivity.

Antibodies to the S proteins of SARS-CoV and SARS-CoV-2—and, to a much lesser extent, MERS-CoV—can cross-react, and both high-potency neutralizing antibodies that also mediate antibody-dependent cytotoxicity and antibody-dependent cellular phagocytosis[69], as well as non-neutralizing antibodies, can be elicited against conserved S epitopes[70,71]. However, the limited spread of SARS-CoV and MERS-CoV means that it is not feasible to assess whether there is any ADE of disease due to SARS-CoV-2 attributable to cross-reactive antibodies[72]. A finding that pre-existing antibodies for other coronaviruses correlate with the low incidence of symptomatic SARS-CoV-2 infection in children would support protection rather than a risk of disease enhancement[73]. To answer this question, the broad application of serological assays that quantify antibodies to virus-specific and cross-reactive epitopes of human coronaviruses in relation to the outcomes of natural infection and of vaccine and antibody trials is required.

The administration of passive antibodies could also reveal whether antibodies predispose to ADE of disease. In small studies, patients infected with SARS or MERS received polyclonal antibodies without apparent worsening of their illness[74–77], and from a meta-analysis it was concluded that early treatment with plasma from patients that had recovered from SARS-CoV infection correlated with a better outcome[76]. In 10 patients with severe COVID-19 that were given plasma with neutralizing titres greater than 1:640 (200 ml) at a median of 16.5 days after disease onset, viraemia was no longer detected and clinical parameters improved within 3 days[78]. Similar findings were reported for 5 severely ill patients treated with plasma with neutralizing titres greater than 1:40[79]; however, another study found no difference in outcome between 52 treated and 51 untreated patients[80]. The evidence that COVID-19 does not worsen after treatment with plasma from convalescent patients has been substantially reinforced by a study of 20,000 patients who were severely ill with the disease, showing an adverse event incidence of 1–3%[81]. If further substantiated, these findings will markedly diminish the concern that clinically relevant amplification of infection, release of immunopathogenic cytokines or immune-complex deposition in the presence of a high viral load is mediated by SARS-CoV-2 antibody-dependent mechanisms[82,83].

High-dose intravenous polyclonal IgG (IVIg)—which is used to treat systemic lupus erythematosus (SLE), idiopathic thrombocytopenia and Kawasaki syndrome[84]—is thought to exert its beneficial effects through the activation of FcγR inhibitory signalling. Because severe COVID-19 could reflect immune dysregulation, a benefit and/or lack of adverse effects in patients receiving plasma from convalescent individuals might reflect the suppression of inflammation induced by IgG, rather than supporting the conclusion that passive antibodies do not trigger ADE of disease through Fab- or Fc-dependent mechanisms. However, the dose of IgG administered to patients with SLE (2 g per kg over 5 days)[85] is much higher than the dose received from convalescent plasma, based on the expected IgG concentrations in plasma (around 500–800 mg per 100 ml) and the amount of convalescent plasma received (200 ml)[78,79]. Assuming a concentration of 1,600 mg per 200 ml, the IgG levels after receiving convalescent plasma (1.6 g per 80 kg) would be approximately 100-fold less than after receiving IVIg (160 g per 80 kg). It is therefore unlikely that the immunomodulatory effects of polyclonal non-antigen-specific IgG dampened possible manifestations of enhanced illness.

Clinically, infections with SARS-CoV, MERS-CoV and SARS-CoV-2 are often biphasic, with more severe respiratory symptoms developing after a week or more and, in some patients, in association with the release of pro-inflammatory cytokines. This pattern has led to the hypothesis that an emerging immune response—including low-avidity, poorly neutralizing antibodies—could exacerbate the disease. However, reports that relate antibody titres to disease progression involve relatively few patients[86–88], and are confounded by the higher levels of antigen seen in severe infections that are predicted to drive a stronger immune response or a heightened innate inflammatory response. One report of three cases of fatal SARS-CoV infection reported that high neutralizing anti-S antibodies and a prominent CD163$^+$ monocyte/macrophage pulmonary infiltrate of cells were associated with reduced expression of TGF-β and CD206$^+$, which are proposed to be markers of macrophages with beneficial functions[89]. However, quantitative analysis of these changes and evidence of an antibody-mediated pathology that is dependent on these cells were not reported. A recent meta-analysis found no relationship between the kinetics of antibody responses to SARS-CoV, MERS-CoV or SARS-CoV-2 and clinical outcomes[90]. At present, there is no evidence that ADE of disease is a factor in the severity of COVID-19. Instead, lung pathology is characterized by diffuse alveolar damage, pneumocyte desquamation, hyaline membranes, neutrophil or macrophage alveolar infiltrates and viral infection of epithelial cells and type II pneumocytes[91]. Further, if instances of ADE of disease occur at all, the experience with dengue suggests that this or other types of immune enhancement will be rare and will occur under highly specific conditions. The aetiology of the inflammatory, Kawasaki-like syndrome that has been associated with SARS-CoV-2 infection in children is unknown, but has not been associated with antibody responses so far[92].

In summary, current clinical experience is insufficient to implicate a role for ADE of disease, or immune enhancement by any other mechanism, in the severity of COVID-19 (Table 1). Prospective studies that relate the kinetics and burden of infection and the host response—including the magnitude, antigen-specificity and molecular mechanisms of action of antibodies, antibody classes and T cell subpopulations—to clinical outcomes are needed to define the characteristics of a beneficial compared with a failed or a potentially detrimental host response to SARS-CoV-2 infection. Although it will probably continue to be difficult to prove that ADE of disease is occurring, or to predict when it might occur, it should be possible to identify correlates of protection that can inform immune-based approaches to the COVID-19 pandemic.

Effects of antibodies on SARS-CoV and MERS-CoV

In vitro studies of the effects of antibodies on viral infection have been used extensively to seek correlates or predictors of ADE of disease (Table 1). These efforts are complicated by the fact that the same antibody mechanisms that are often proposed to result in ADE of infection are responsible for protection from viral disease in vivo. Although infection was most often blocked by anti-S antibodies, several reports have shown antibody-dependent uptake of SARS-CoV or SARS-CoV S-pseudoviruses that was mediated by binding of the Fab component to the virus and the Fc component to FcγR on the target cell (Fig. 2) using in vitro methods[93–98]. Importantly, viral uptake did not result in productive infection. An antibody that binds the S protein and mimics receptor-mediated entry to facilitate viral uptake has been described for MERS-CoV[99], but not for SARS-CoV or SARS-CoV-2. Although SARS-CoV and SARS-CoV-2 do not infect myeloid cells[100–103], the productive infection of macrophages by MERS-CoV has been reported, albeit at low levels[104]. It is notable that higher production of immune-cell-attracting chemokines was observed in myeloid cells infected by MERS-CoV but not in cells exposed to SARS-CoV, suggesting that productive infection has a greater effect on this response[104]. The biology of the interactions of coronaviruses with cells expressing FcγRs is therefore very different from the targeting of FcγR-expressing myeloid cells by the dengue viruses. Conversely, in vitro methods can reliably define the properties of mAbs or of vaccine-induced antibodies—including their epitope specificity, binding affinity and avidity, and maturation as well as any potential to enhance fusion, together with their capacities for neutralization and antiviral Fc-dependent effector functions (Fig. 2).

Antibody effects in coronavirus-infected animals

Small-animal models

Several mouse, rat and other small-animal models of SARS-CoV infection have used passive-antibody administration or immunization to investigate whether pre-existing antibodies protect against or enhance disease. Although vaccine enhancement of disease in these models could occur through other mechanisms, such studies can directly assess the protective or enhancing properties of passive antibodies (Table 1).

In the ferret model of SARS-CoV infection, a human mAb was found to protect the animals from infection[105]; however, modified vaccinia Ankara expressing S protein (MVA-S) was not protective and liver inflammation was noted in this model[106]. Pre- and post-exposure administration of a mAb against MERS-CoV protected mice from challenge, as assessed by lung viral load, lung pathology and weight loss[107]. Three mAbs against SARS-CoV, given at a high dose before challenge, protected young and old mice against lung viral spread and inflammation, but had no effect when given after infection[108]. Low doses were less protective, but no ADE of disease was observed. A caveat is that human mAbs were tested in the context of mouse FcγRs; however, this can be addressed using human FcgR transgenic animals[109]. Both previous infection and passive transfer of mouse neutralizing antibodies partially protected 4–6-week-old mice against secondary infection with SARS-CoV[110], and no ADE of disease was observed despite low neutralizing titres. In another mouse study[111], passive transfer of SARS-CoV-immune serum was found to mediate protection by Fc-dependent monocyte effector function through antibody-dependent cellular phagocytosis; however, natural killer cells, antibody-dependent cytotoxicity or complement-antibody complexes did not contribute to protection. In a mouse model of vaccination, which used SARS-CoV in which the E protein had been deleted as a live attenuated vaccine, induction of antibodies and T cell immunity and protection against lethal viral challenge was observed in mice from three age groups[112]. By contrast, enhanced disease was observed in mice that were immunized with formalin- or ultraviolet-inactivated SARS-CoV. Whereas younger mice were protected, older mice developed pulmonary pathology with an eosinophil infiltrate; this suggests a detrimental Th2 response related to age, rather than ADE of disease[113]. In some models, cellular immunopathology might be linked to Th17-mediated activation of eosinophils[114]. In another report, mice given formalin- or ultraviolet-inactivated SARS-CoV or other vaccine formulations developed neutralizing antibodies and were protected from challenge, but also developed eosinophilic pulmonary infiltrates[115]. This type of immunopathology has not been reported in fatal human coronavirus infections.

Small-animal studies of SARS-CoV-2 infection are being reported rapidly. Neutralizing antibodies to SARS-CoV-2 were induced by immunizing rats with the RBD of the S protein and adjuvant[94]. In vitro evaluation of the potential for enhanced uptake of SARS-CoV-2 using HEK293T cells expressing rat FcγRI in the presence or absence of ACE2 expression showed neutralization but no enhancement of viral entry. Mice that were given an mRNA vaccine expressing pre-fusion SARS-CoV-2 S protein developed neutralizing antibodies and S-protein-specific CD8 T cell responses that were protective against lung infection without evidence of immunopathology[116], and neutralizing mAbs against the RBD of the S protein of SARS-CoV-2 reduced lung infection and cytokine release[117].

Table 1 | Information provided by and limitations of approaches for the assessment of antibody-mediated protection against SARS-CoV-2 and the potential for antibody-dependent enhancement of disease

Test modality	Information provided	Limitations
In vitro: cell culture Infect relevant human cells with or without antibodies	Virus neutralization Virus uptake, productive infection or cytokines	Cell lines lack primary cell receptor characteristics Primary human cells are difficult to culture and have donor variability • Receptor expression must be maintained
In vivo: animal models Infection of animals with or without antibody or vaccine intervention	Protection against or increase of viral replication or disease	Lack of disease models of human illness Lack of models predictive of enhanced disease in humans Viral replication as a proxy of disease requires clinical validation Need to assess T cells for contribution to pathology or reducing ADE With human mAbs: • Differential engagement of animal FcγRs • Different expression patterns of FcγRs in humans and animals • Potential generation of anti-human antibodies
Human: clinical and epidemiological studies	Correlations of outcomes with • Previous HCoV infection • Treatment with plasma from convalescent patients • Kinetics of adaptive immune responses	No markers to differentiate severe disease from enhanced disease Limited knowledge of antibody or T cell epitope specificities during natural SARS-CoV-2 or other HCoV infection, and of outcomes of infection with new coronaviruses

Passive transfer of a neutralizing antibody protected Syrian hamsters against high-dose SARS-CoV-2, as demonstrated by maintained weight and low lung viral titres[118]. Similarly, hamsters immunized with recombinant SARS-CoV S protein trimer developed neutralizing antibodies and were protected against challenge[119]. Whereas serum from vaccinated hamsters mediated FcγRIIb-dependent enhancement of SARS-CoV entry into B cell lines, virus replication was abortive in vitro and viral load and lung pathology were not increased in vaccinated animals[98]. These data underscore that enhancement of viral entry into cells in vitro does not predict negative consequences in vivo, further highlighting the important gap between in vitro findings and the causes of ADE of disease in vivo.

Unlike SARS-CoV, MERS-CoV and SARS-CoV-2, feline infectious peritonitis virus is an alphacoronavirus that, as with dengue, has tropism for macrophages. Infection with this virus has been shown to be enhanced by pre-existing antibodies, especially those against the same strain[120].

Non-human primate models

In non-human primates (NHPs), infection with SARS-CoV, MERS-CoV or SARS-CoV-2 results in viral spread to multiple tissues, including lungs[121–123]. Rhesus macaques that were administered a high inoculum of SARS-CoV-2 by nasal, tracheal, ocular and oral routes had increased temperatures and respiratory rates for 1 day, and reduced appetite and dehydration for 9–16 days[122]. Macaques that were euthanized at 3 days and 21 days had multifocal lung lesions, with alveolar septal thickening due to oedema and fibrin, small to moderate numbers of macrophages, a few neutrophils, minimal type II pneumocyte hyperplasia and some perivascular lymphocyte cuffing. SARS-CoV-2 viral proteins were detected in a few type I and type II pneumocytes, and alveolar macrophages and virions were found in type I pneumocytes. Although these foci of lung pathology have some similarities to those observed in human infection[91], NHPs develop minimal or no signs of respiratory or systemic betacoronavirus disease.

After the outbreaks of SARS-CoV and MERS-CoV disease, NHPs were used in the evaluation of several vaccine and antibody interventions (Supplementary Table 1). In one study, FI-SARS-CoV reduced viraemia and protected against lung pathology in rhesus macaques[124], whereas in another study macaques given FI-SARS-CoV developed macrophage and lymphocytic infiltrates and alveolar oedema with fibrin deposition after challenge, indicating the difficulties of establishing consistent NHP models[125]. Synthetic peptide vaccines have also been prepared using sera from convalescent patients to define immunodominant epitopes of SARS-CoV S protein[125]. The vaccines were found to reduce pathology after SARS-CoV challenge unless the S protein of the vaccine included amino acids 597–603, suggesting an epitope-specific basis for the induction of lung pathology. However, these peptide constructs would not be expected to fully mimic antibody or T cell responses that would be elicited to the intact S protein.

Two studies have reported the immunization of rhesus macaques with MVA expressing SARS-CoV S protein or an MVA control. In the first report, three out of four immunized macaques had no detectable shedding or enhanced lung infection 7 days after challenge[126]. In the second report, immunization elicited polyclonal anti-S antibodies with neutralizing activity and reduced infection in three out of eight macaques after challenge[89]. However, although the challenge inoculum was the same as in the first study, areas of diffuse alveolar damage were detected in six out of eight vaccinated macaques compared with one out of eight control animals euthanized at 7 days, as well as at 35 days. Immunization with MVA-S was associated with an accumulation of monocytes and macrophages, and with the detection of activated alveolar macrophages that produced pro-inflammatory MCP-1 and IL-8, which were were not observed in control animals. In a second cohort that was given polyclonal IgG from vaccinated macaques or control animals, loss of TGF-β and increased IL-6 production by activated pulmonary macrophages was observed in macaques that were pre-treated with anti-S IgG, and lung pathology was described as skewed towards immunopathological inflammation. However, it was not stated whether the histopathology was focal or widespread in the lungs, and immunopathology was not associated with impaired respiratory function in macaques evaluated for 21 days (passive anti-S) or for 35 days (MVA-S). Although differences in macrophage markers were associated with changes in the lungs, a causal relationship between anti-S antibodies and an antibody-dependent macrophage-mediated mechanism of more severe pathological changes was not explored, and whether MVA-S might have generated non-neutralizing antibodies that enhanced lung pathology was not assessed. It will therefore be important to define the epitope specificity and serum neutralization activity in these animal models, and potential T cell mechanisms will need to be excluded before enhanced immunopathology can be attributed to antibody mechanisms.

The second study reporting immunization of rhesus macaques with MVA-S[89] also described in vitro experiments using sera from patients who had recovered from SARS-CoV infection. However, only one out of eight sera samples elicited enhanced cytokine production by human macrophages in vitro. Because IL-8 production by macrophages treated with one of the serum samples was lower in the presence of FcγR-blocking antibody (no control serum), it was concluded that blocking FcγRs might be necessary to reduce lung damage caused by SARS-CoV. However, the finding was not confirmed with sera from other severe cases of SARS, and is subject to the caveat that in vitro studies cannot be taken as evidence of ADE of disease.

In contrast to the immunopathology observed after immunization with MVA-S, other studies of SARS-CoV have suggested a protective effect of vaccine-induced antibodies. Using a purified SARS-CoV-infected cell lysate as a vaccine, cynomolgus macaques were protected from challenge, and low neutralizing antibody titres were not associated with ADE of disease[127]. Further, African green monkeys with pre-existing antibody and/or T cells after primary SARS-CoV infection were protected from homologous re-challenge as assessed by lung virus titres, although the pulmonary inflammatory response was not different from that of primary infection[128].

In additional studies, rhesus macaques immunized with a chimpanzee adenovirus (ChAdOx1 MERS) expressing MERS-CoV S protein, a recombinant S-RBD protein or a synthetic MERS-CoV S DNA vaccine, had decreased infection and no enhanced lung pathology upon challenge[129–131].

The potential for immune enhancement of SARS-CoV-2 infection by antibody-dependent or other mechanisms has been assessed by infection and re-challenge of rhesus macaques. Out of two rhesus macaques that were re-challenged 28 days after initial infection—when neutralizing antibody titres were low (1:8–1:16)—neither exhibited viral shedding and one had no lung pathology. Immunity to SARS-CoV-2 in nine rhesus macaques—including the presence of neutralizing antibodies, antibody-mediated effector functions and antiviral CD4 and CD8 T cells—was associated with protection upon re-challenge at 35 days[123]. When vaccines were tested, rhesus macaques immunized with purified β-propriolactone-inactivated SARS-CoV-2 in alum showed complete or partial protection against high-inoculum SARS-CoV-2 challenge, and histopathological analyses of lungs and other organs at 29 days showed no evidence of ADE of disease compared with control macaques[132]. A large study involving 35 rhesus macaques, which were given prototype DNA vaccines expressing either full-length SARS-CoV-2 S protein or components of this protein, found that protection was correlated with the presence of neutralizing antibodies—and, notably, with Fc-dependent antibody effector functions—and there were no adverse outcomes after challenge[133].

In studies of neutralizing mAbs (Supplementary Table 1), viral titres and lung pathology after nasal challenge were reduced in rhesus macaques that were administered a mAb directed against a proteolytic cleavage site in the SARS-CoV S protein that is required for host-cell entry[134]. Macaques given mAbs against MERS-CoV showed less pulmonary involvement and no worsening of disease with challenge[135]. The prophylactic administration of mAbs against MERS-CoV to marmosets one day before challenge was associated with reduced lung pathology compared with the administration of control mAbs[136–138]; mAbs were found to be protective when administered 2–12 h after challenge but not when given 1 day after challenge[137,138]. These animal studies of coronavirus infections parallel the observation that the passive transfer of mAbs against RSV that have selected properties can be protective, whereas a particular vaccine formulation (FI-RSV) that is directed to the same viral protein can enhance disease.

In summary, in most animal models—including NHPs—vaccination or the administration of passive mAbs have demonstrated protection against challenge with SARS-CoV, MERS-CoV or SARS-CoV-2, although reports on SARS-CoV-2 are limited. However, studies of an FI-SARS-CoV vaccine, one of two studies of an MVA vaccine expressing SARS-CoV S protein, and vaccination with one S-derived peptide showed enhanced lung pathology in NHPs. Thus, there are limited data to indicate that immune responses that include antibodies (and probably also T cells) induced by some vaccine formulations may be associated with more extensive lung pathology compared with infection alone, whereas the transfer of mAbs with specific properties have, so far, provided protection in animals (Supplementary Table 1).

Overall, the lack of a link between clinical measures of disease severity in NHPs and the experimental conditions associated with exacerbated lung pathology is a limitation to their utility in predicting the risks of ADE associated with passive-antibody or vaccine interventions in humans. So far, the models do not emulate the severe respiratory disease observed in COVID-19. Evaluation of T cell responses will also be needed to draw conclusions regarding mechanisms if immunopathology is observed. For example, a strong T cell response has been described as ameliorating ADE of disease in a dengue model[139] and animal studies have suggested an aberrant T cell response to FI-RSV vaccination[33,114]. Quantitative assessments of the extent of lung involvement, and histopathological scoring of the characteristics and severity of lesions using validated markers of infected cells, patterns of cell-subtype infection and quantification of infiltrating immune cells will be also be necessary before these models can be used to better understand either protective immunity or immune enhancement—whether mediated by antibodies, T cells, intrinsic responses or a combination of factors. A critical point is that the identification of correlates of protection in humans will be necessary to understand how studies in small- and large-animal models can be designed to support or question the benefits of particular immune interventions for SARS-CoV-2 infection.

Conclusions

It is clear that after many years, and considerable attention, the understanding of ADE of disease after either vaccination or administration of antiviral antibodies is insufficient to confidently predict that a given immune intervention for a viral infection will have negative outcomes in humans. Despite the importance that such information would have in the COVID-19 pandemic, in vitro assays do not predict ADE of disease. Most animal models of vaccines and antibody interventions show protection, whereas those that suggest potential ADE of disease are not definitive and the precise mechanisms have not been defined. Although ADE is a concern, it is also clear that antibodies are a fundamentally important component of protective immunity to all of the pathogens discussed here, and that their protective effects depend both on the binding of viral proteins by their Fab fragments and on the effector functions conferred by their Fc fragments. Even when vaccine formulations such as formalin inactivation have shown disease enhancement, neutralizing antibodies with optimized properties have been protective. Further, the potential mechanisms of ADE of disease are probably virus-specific and, importantly, clinical markers do not differentiate severe infection from immune enhancement. Additional mechanism-focused studies are needed to determine whether small-animal and NHP models of virus infection, including for SARS-CoV-2, can predict the probable benefits or risks of vaccines or passive-antibody interventions in humans. Optimizing these models must be informed by understanding the correlates of protection against SARS-CoV-2 in natural human infection and as vaccines and antibodies are evaluated in humans. Such mechanistic and in vivo studies across viral pathogens are essential so that we are better prepared to face future pandemics. In the meantime, it will be necessary to directly test safety and define correlates of protection conferred by vaccines and antibodies against SARS-CoV-2 and other viral pathogens in human clinical trials.

1. Luke, T. C., Kilbane, E. M., Jackson, J. L. & Hoffman, S. L. Meta-analysis: convalescent blood products for Spanish influenza pneumonia: a future H5N1 treatment? *Ann. Intern. Med.* **145**, 599–609 (2006).
2. Casadevall, A., Dadachova, E. & Pirofski, L. A. Passive antibody therapy for infectious diseases. *Nat. Rev. Microbiol.* **2**, 695–703 (2004).
3. Plotkin, S. A. Correlates of protection induced by vaccination. *Clin. Vaccine Immunol.* **17**, 1055–1065 (2010).
4. VanBlargan, L. A., Goo, L. & Pierson, T. C. Deconstructing the antiviral neutralizing-antibody response: implications for vaccine development and immunity. *Microbiol. Mol. Biol. Rev.* **80**, 989–1010 (2016).
5. Corti, D. & Lanzavecchia, A. Broadly neutralizing antiviral antibodies. *Ann. Rev. Immunol.* **31**, 705–742 (2013).
6. Walker, L. M. & Burton, D. R. Passive immunotherapy of viral infections: 'super-antibodies' enter the fray. *Nat. Rev. Immunol.* **18**, 297–308 (2018).
7. Lu, L. L., Suscovich, T. J., Fortune, S. M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46–61 (2018).
8. Bournazos, S. & Ravetch, J. V. Fcγ receptor function and the design of vaccination strategies. *Immunity* **47**, 224–233 (2017).
9. DiLillo, D. J., Tan, G. S., Palese, P. & Ravetch, J. V. Broadly neutralizing hemagglutinin stalk-specific antibodies require FcγR interactions for protection against influenza virus in vivo. *Nat. Med.* **20**, 143–151 (2014).
10. Bournazos, S. et al. Broadly neutralizing anti-HIV-1 antibodies require Fc effector functions for in vivo activity. *Cell* **158**, 1243–1253 (2014).
11. Pyzik, M. et al. The neonatal Fc receptor (FcRn): a misnomer? *Front. Immunol.* **10**, 1540 (2019).
12. Bergtold, A., Desai, D. D., Gavhane, A. & Clynes, R. Cell surface recycling of internalized antigen permits dendritic cell priming of B cells. *Immunity* **23**, 503–514 (2005).
13. Nishimura, Y. et al. Early antibody therapy can induce long-lasting immunity to SHIV. *Nature* **543**, 559–563 (2017).
14. Gunn, B. M. et al. A Role for Fc function in therapeutic monoclonal antibody-mediated protection against Ebola virus. *Cell Host Microbe* **24**, 221–233.e5 (2018).
15. Graham, B. S. Rapid COVID-19 vaccine development. *Science* **368**, 945–946 (2020).
16. Kim, H. W. et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *Am. J. Epidemiol.* **89**, 422–434 (1969).
17. Kapikian, A. Z., Mitchell, R. H., Chanock, R. M., Shvedoff, R. A. & Stewart, C. E. An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine. *Am. J. Epidemiol.* **89**, 405–421 (1969).
18. Polack, F. P., Hoffman, S. J., Crujeiras, G. & Griffin, D. E. A role for nonprotective complement-fixing antibodies with low avidity for measles virus in atypical measles. *Nat. Med.* **9**, 1209–1213 (2003).
19. Simmons, C. P., Farrar, J. J., Nguyen, V. & Wills, B. Dengue. *N. Engl. J. Med.* **366**, 1423–1432 (2012).
20. Katzelnick, L. C. et al. Antibody-dependent enhancement of severe dengue disease in humans. *Science* **358**, 929–932 (2017).
21. Guzman, M. G., Alvarez, M. & Halstead, S. B. Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection. *Arch. Virol.* **158**, 1445–1459 (2013).
22. Iwasaki, A. & Yang, Y. The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev. Immunol.* **20**, 339–341 (2020).
23. Dekkers, G. et al. Affinity of human IgG subclasses to mouse Fc gamma receptors. *MAbs* **9**, 767–773 (2017).
24. Crowley, A. R. & Ackerman, M. E. Mind the gap: how interspecies variability in IgG and its receptors may complicate comparisons of human and non-human primate effector function. *Front. Immunol.* **10**, 697 (2019).
25. Fulginiti, V. A. et al. Respiratory virus immunization. A field trial of two inactivated respiratory virus vaccines; an aqueous trivalent parainfluenza virus vaccine and an alum-precipitated respiratory syncytial virus vaccine. *Am. J. Epidemiol.* **89**, 435–448 (1969).
26. Chin, J., Magoffin, R. L., Shearer, L. A., Schieble, J. H. & Lennette, E. H. Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population. *Am. J. Epidemiol.* **89**, 449–463 (1969).
27. Murphy, B. R. et al. Dissociation between serum neutralizing and glycoprotein antibody responses of infants and children who received inactivated respiratory syncytial virus vaccine. *J. Clin. Microbiol.* **24**, 197–202 (1986).
28. Polack, F. P. et al. A role for immune complexes in enhanced respiratory syncytial virus disease. *J. Exp. Med.* **196**, 859–865 (2002).
29. Atkinson, J. P. et al. The human complement system: basic concepts and clinical relevance. *Clin. Immunol.* https://doi.org/10.1016/B978-0-7020-6896-6.00021-1 (2019).
30. Kim, H. W. et al. Cell-mediated immunity to respiratory syncytial virus induced by inactivated vaccine or by infection. *Pediatr. Res.* **10**, 75–78 (1976).
31. van Erp, E. A., Luytjes, W., Ferwerda, G. & van Kasteren, P. B. Fc-mediated antibody effector functions during respiratory syncytial virus infection and disease. *Front. Immunol.* **10**, 548 (2019).
32. Delgado, M. F. et al. Lack of antibody affinity maturation due to poor Toll-like receptor stimulation leads to enhanced respiratory syncytial virus disease. *Nat. Med.* **15**, 34–41 (2009).
33. Ruckwardt, T. J., Morabito, K. M. & Graham, B. S. Immunological lessons from respiratory syncytial virus vaccine development. *Immunity* **51**, 429–442 (2019).
34. Aranda, S. S. & Polack, F. P. Prevention of pediatric respiratory syncytial virus lower respiratory tract illness: perspectives for the next decade. *Front. Immunol.* **10**, 1006 (2019).
35. Regeneron to discontinue development of Suptavumab for respiratory syncytial virus. https://investor.regeneron.com/news-releases/news-release-details/regeneron-discontinue-development-suptavumab-respiratory (2017).
36. Domachowske, J. B. et al. Safety, tolerability and pharmacokinetics of MEDI8897, an extended half-life single-dose respiratory syncytial virus prefusion F-targeting monoclonal antibody administered as a single dose to healthy preterm infants. *Pediatr. Infect. Dis. J.* **37**, 886–892 (2018).
37. Ng, S. et al. Novel correlates of protection against pandemic H1N1 influenza A virus infection. *Nat. Med.* **25**, 962–967 (2019).
38. Skowronski, D. M. et al. Association between the 2008–09 seasonal influenza vaccine and pandemic H1N1 illness during spring–summer 2009: four observational studies from Canada. *PLoS Med.* **7**, e1000258 (2010).

39. Wu, J. T. et al. The infection attack rate and severity of 2009 pandemic H1N1 influenza in Hong Kong. *Clin. Infect. Dis*. **51**, 1184–1191 (2010).
40. Lansbury, L. E. et al. Effectiveness of 2009 pandemic influenza A(H1N1) vaccines: a systematic review and meta-analysis. *Vaccine* **35**, 1996–2006 (2017).
41. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis*. **12**, 36–44 (2012).
42. Monsalvo, A. C. et al. Severe pandemic 2009 H1N1 influenza disease due to pathogenic immune complexes. *Nat. Med*. **17**, 195–199 (2011).
43. Co, M. D. T. et al. Relationship of preexisting influenza hemagglutination inhibition, complement-dependent lytic, and antibody-dependent cellular cytotoxicity antibodies to the development of clinical illness in a prospective study of A(H1N1)pdm09 influenza in children. *Viral Immunol*. **27**, 375–382 (2014).
44. Khurana, S. et al. Vaccine-induced anti-HA2 antibodies promote virus fusion and enhance influenza virus respiratory disease. *Sci. Transl. Med*. **5**, 200ra114 (2013).
45. Winarski, K. L. et al. Antibody-dependent enhancement of influenza disease promoted by increase in hemagglutinin stem flexibility and virus fusion kinetics. *Proc. Natl Acad. Sci. USA* **116**, 15194–15199 (2019).
46. Beltramello, M. et al. The human immune response to dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. *Cell Host Microbe* **8**, 271–283 (2010).
47. de Alwis, R. et al. Dengue viruses are enhanced by distinct populations of serotype cross-reactive antibodies in human immune sera. *PLoS Pathog*. **10**, e1004386 (2014).
48. Thomas, S. J. & Yoon, I.-K. A review of Dengvaxia®: development to deployment. *Hum. Vaccin. Immunother*. **15**, 2295–2314 (2019).
49. WHO Report. Dengue vaccine: WHO position paper, September 2018 – Recommendations. *Vaccine* **37**, 4848–4849 (2019).
50. Rodriguez-Barraquer, I. et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. *Science* **363**, 607–610 (2019).
51. Chan, K. R. et al. Cross-reactive antibodies enhance live attenuated virus infection for increased immunogenicity. *Nat. Microbiol*. **1**, 16164 (2016).
52. Browne, S. K., Beeler, J. A. & Roberts, J. N. Summary of the vaccines and related biological products advisory committee meeting held to consider evaluation of vaccine candidates for the prevention of respiratory syncytial virus disease in RSV-naïve infants. *Vaccine* **38**, 101–106 (2020).
53. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
54. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e6 (2020).
55. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
56. Muus, C. et al. Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. Preprint at https://www.biorxiv.org/content/10.1101/2020.04.19.049254v2 (2020).
57. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med*. **26**, 681–687 (2020).
58. Ziegler, C. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is enriched in specific cell subsets across tissues. *Cell* **181**, 1016–1035.e19 (2020).
59. Wellcome Sanger Institute, Human Cell Atlas & Chan Zuckerberg Initiative. *COVID-19 Cell Atlas*, https://www.covid19cellatlas.org/
60. Chan Zuckerberg Biohub & Stanford University. *Lung Cell Atlas*, https://hlca.ds.czbiohub.org/
61. Ng, K. et al. Pre-existing and de novo humoral immunity to SARS-CoV-2 in humans. Preprint at https://www.biorxiv.org/content/10.1101/2020.05.14.095414v1 (2020).
62. Braun, J. et al. Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and healthy donors. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.17.20061440v1 (2020).
    Detection of anti-S protein CD4$^+$ T cells in 83% patients with COVID-19 with reactivity to epitopes in both N- and C-terminal domains, and in 34% of healthy unexposed donors, indicating cross-reactive T cell immunity against SARS-CoV-2 attributable to previous coronavirus infections, with epitopes predominantly in the C-terminal domain that has higher homology to other coronaviruses.
63. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501.e15 (2020).
    Extensive analysis of CD4 and CD8 T cell responses to epitopes of S-, M- and N proteins as well as non-structural proteins of SARS-CoV-2 in convalescent patients with COVID-19 and detection of cross-reactive CD4$^+$ T cells that recognized SAR-CoV-2 epitopes in 40–60% of unexposed donors.
64. van der Hoek, L., Pyrc, K. & Berkhout, B. Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol. Rev*. **30**, 760–773 (2006).
65. Callow, K. A., Parry, H. F., Sergeant, M. & Tyrrell, D. A. J. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol. Infect*. **105**, 435–446 (1990).
66. Reed, S. E. The behaviour of recent isolates of human respiratory coronavirus in vitro and in volunteers: evidence of heterogeneity among 229E-related strains. *J. Med. Virol*. **13**, 179–192 (1984).
67. Chan, K. H. et al. Serological responses in patients with severe acute respiratory syndrome coronavirus infection and cross-reactivity with human coronaviruses 229E, OC43, and NL63. *Clin. Diagn. Lab. Immunol*. **12**, 1317–1321 (2005).
68. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
69. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
70. Okba, N. M. A. et al. Severe acute respiratory syndrome coronavirus 2–specific antibody responses in coronavirus disease patients. *Emerging Infect. Dis*. **26**, 1478–1488 (2020).
71. Lv, H. et al. Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections. *Cell Reports* **31**, 107725 (2020).
72. Guo, X. et al. Long-term persistence of IgG antibodies in SARS-CoV infected healthcare workers. Preprint at https://www.medrxiv.org/content/10.1101/2020.02.12.20021386v1 (2020).
73. Lavezzo, E. et al. Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *Nature* https://doi.org/10.1038/s41586-020-2488-1 (2020).
74. Yeh, K.-M. et al. Experience of using convalescent plasma for severe acute respiratory syndrome among healthcare workers in a Taiwan hospital. *J. Antimicrob. Chemother*. **56**, 919–922 (2005).
75. Cheng, Y. et al. Use of convalescent plasma therapy in SARS patients in Hong Kong. *Eur. J. Clin. Microbiol. Infect. Dis*. **24**, 44–46 (2005).
76. Mair-Jenkins, J. et al. The effectiveness of convalescent plasma and hyperimmune immunoglobulin for the treatment of severe acute respiratory infections of viral etiology: a systematic review and exploratory meta-analysis. *J. Infect. Dis*. **211**, 80–90 (2015).
77. Ko, J.-H. et al. Challenges of convalescent plasma infusion therapy in Middle East respiratory coronavirus infection: a single centre experience. *Antivir. Ther*. **23**, 617–622 (2018).
78. Duan, K. et al. Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *Proc. Natl Acad. Sci. USA* **117**, 9490–9496 (2020).
79. Shen, C. et al. Treatment of 5 critically ill patients with COVID-19 with convalescent plasma. *J. Am. Med. Assoc*. **323**, 1582–1589 (2020).
80. Li, L. et al. Effect of convalescent plasma therapy on time to clinical improvement in patients with severe and life-threatening COVID-19: a randomized clinical trial. *J. Am. Med. Assoc*. https://doi.org/10.1001/jama.2020.10044 (2020).
81. Joyner, M. J & Wright, R. S. Safety update: COVID-19 convalescent plasma in 20,000 hospitalized patients. *Mayo Clin. Proc*. https://doi.org/10.1016/j.mayocp.2020.06.028 (2020).
    Major US-wide study of the administration of convalescent plasma to patients with COVID-19 with severe respiratory disease, followed by observation for seven days post-infusion with no evidence of disease progression associated with passive-antibody therapy.
82. Chen, L., Xiong, J., Bao, L. & Shi, Y. Convalescent plasma as a potential therapy for COVID-19. *Lancet Infect. Dis*. **20**, 398–400 (2020).
83. de Alwis, R., Chen, S., Gan, E. S. & Ooi, E. E. Impact of immune enhancement on COVID-19 polyclonal hyperimmune globulin therapy and vaccine development. *EBioMedicine* **55**, 102768 (2020).
84. Galeotti, C., Kaveri, S. V. & Bayry, J. IVIG-mediated effector functions in autoimmune and inflammatory diseases. *Int. Immunol*. **29**, 491–498 (2017).
85. Zandman-Goddard, G., Levy, Y. & Shoenfeld, Y. Intravenous immunoglobulin therapy and systemic lupus erythematosus. *Clin. Rev. Allergy Immunol*. **29**, 219–228 (2005).
86. Lee, N. et al. Anti-SARS-CoV IgG response in relation to disease severity of severe acute respiratory syndrome. *J. Clin. Virol*. **35**, 179–184 (2006).
87. Zhang, L. et al. Antibody responses against SARS coronavirus are correlated with disease outcome of infected individuals. *J. Med. Virol*. **78**, 1–8 (2006).
88. Ho, M.-S. et al. Neutralizing antibody response and SARS severity. *Emerg. Infect. Dis*. **11**, 1730–1737 (2005).
89. Liu, L. et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* **4**, e123158 (2019).
90. Huang, A. et al. A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of disease. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.14.20065771v1 (2020).
    Meta-analysis of reports of antibody responses to SARS-CoV, MERS-CoV and initial reports of SARS-CoV-2 in infected patients, describing inconclusive evidence for a relationship between antibody titres and disease severity.
91. Martines, R. B. et al. Pathology and pathogenesis of SARS-CoV-2 associated with fatal coronavirus disease, United States. *Emerg. Infect. Dis*. https://doi.org/10.3201/eid2609.202095 (2020).
92. Ramcharan, T. et al. Paediatric inflammatory multisystem syndrome: temporally associated with SARS-CoV-2 (PIMS-TS): cardiac features, management and short-term outcomes at a UK tertiary paediatric hospital. *Pediatr. Cardiol*. https://doi.org/10.1007/s00246-020-02391-2 (2020).
93. Yang, Z.-Y. et al. pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J. Virol*. **78**, 5642–5650 (2004).
94. Quinlan, B. D. et al. The SARS-CoV-2 receptor-binding domain elicits a potent neutralizing response without antibody-dependent enhancement. Preprint at https://www.biorxiv.org/content/10.1101/2020.04.10.036418v1 (2020).
95. Yip, M. S. et al. Antibody-dependent enhancement of SARS coronavirus infection and its role in the pathogenesis of SARS. *Hong Kong Med. J*. **22**, 25–31 (2016).
96. Yip, M. S. et al. Antibody-dependent infection of human macrophages by severe acute respiratory syndrome coronavirus. *Virol. J*. **11**, 82 (2014).
97. Wang, S.-F. et al. Antibody-dependent SARS coronavirus infection is mediated by antibodies against spike proteins. *Biochem. Biophys. Res. Commun*. **451**, 208–214 (2014).
98. Jaume, M. et al. Anti-severe acute respiratory syndrome coronavirus spike antibodies trigger infection of human immune cells via a pH- and cysteine protease-independent FcγR pathway. *J. Virol*. **85**, 10582–10597 (2011).
99. Wan, Y. et al. Molecular mechanism for antibody-dependent enhancement of coronavirus entry. *J. Virol*. **94**, e02015–e02019 (2020).
100. Yilla, M. et al. SARS-coronavirus replication in human peripheral monocytes/macrophages. *Virus Res*. **107**, 93–101 (2005).
101. Lau, Y. L., Peiris, J. S. M. & Law, H. K. W. Role of dendritic cells in SARS coronavirus infection. *Hong Kong Med. J*. **18**, 28–30 (2012).

102. Tynell, J. et al. Middle East respiratory syndrome coronavirus shows poor replication but significant induction of antiviral responses in human monocyte-derived macrophages and dendritic cells. *J. Gen. Virol*. **97**, 344–355 (2016).
103. Hui, K. P. Y. et al. Tropism, replication competence, and innate immune responses of the coronavirus SARS-CoV-2 in human respiratory tract and conjunctiva: an analysis in ex-vivo and in-vitro cultures. *Lancet Respir. Med*. **8**, 687–695 (2020).
104. Zhou, J. et al. Active replication of Middle East respiratory syndrome coronavirus and aberrant induction of inflammatory cytokines and chemokines in human macrophages: implications for pathogenesis. *J. Infect. Dis*. **209**, 1331–1342 (2014).
105. ter Meulen, J. et al. Human monoclonal antibody as prophylaxis for SARS coronavirus infection in ferrets. *Lancet* **363**, 2139–2141 (2004).
106. Czub, M., Weingartl, H., Czub, S., He, R. & Cao, J. Evaluation of modified vaccinia virus Ankara based recombinant SARS vaccine in ferrets. *Vaccine* **23**, 2273–2279 (2005).
107. Corti, D. et al. Prophylactic and postexposure efficacy of a potent human monoclonal antibody against MERS coronavirus. *Proc. Natl Acad. Sci. USA* **112**, 10473–10478 (2015).
108. Rockx, B. et al. Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. *J. Virol*. **82**, 3220–3235 (2008).
109. Smith, P., DiLillo, D. J., Bournazos, S., Li, F. & Ravetch, J. V. Mouse model recapitulating human Fcγ receptor structural and functional diversity. *Proc. Natl Acad. Sci. USA* **109**, 6181–6186 (2012).
110. Subbarao, K. et al. Prior infection and passive transfer of neutralizing antibody prevent replication of severe acute respiratory syndrome coronavirus in the respiratory tract of mice. *J. Virol*. **78**, 3572–3577 (2004).
111. Yasui, F. et al. Phagocytic cells contribute to the antibody-mediated elimination of pulmonary-infected SARS coronavirus. *Virology* **454–455**, 157–168 (2014).
112. Fett, C., DeDiego, M. L., Regla-Nava, J. A., Enjuanes, L. & Perlman, S. Complete protection against severe acute respiratory syndrome coronavirus-mediated lethal respiratory disease in aged mice by immunization with a mouse-adapted virus lacking E protein. *J. Virol*. **87**, 6551–6559 (2013).
113. Bolles, M. et al. A double-inactivated severe acute respiratory syndrome coronavirus vaccine provides incomplete protection in mice and induces increased eosinophilic proinflammatory pulmonary response upon challenge. *J. Virol*. **85**, 12201–12215 (2011).
114. Hotez, P. J., Corry, D. B. & Bottazzi, M. E. COVID-19 vaccine design: the Janus face of immune enhancement. *Nat. Rev. Immunol*. **20**, 347–348 (2020).
115. Tseng, C.-T. et al. Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS ONE* **7**, e35421 (2012).
116. Corbett, K. S. et al. SARS-CoV-2 mRNA vaccine development enabled by prototype pathogen preparedness. Preprint at https://www.biorxiv.org/content/10.1101/2020.06.11.145920v1 (2020).
117. Zost, S. J. et al. Potently neutralizing human antibodies that block SARS-CoV-2 receptor binding and protect animals. Preprint at https://www.biorxiv.org/content/10.1101/2020.05.22.111005v1 (2020).
Protection of mice against SARS-CoV-2 by human mAbs targeting distinct epitopes of the S protein, some of which had synergistic effects in vitro, without evidence of ADE of disease in the animal model.
118. Rogers, T. F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* https://doi.org/10.1126/science.abc7520 (2020).
Protective effects of neutralizing mAbs against RBD and non-RBD epitopes of SARS-CoV-2 S protein without evidence of ADE of disease in a Syrian hamster model.
119. Kam, Y. W. et al. Antibodies against trimeric S glycoprotein protect hamsters against SARS-CoV challenge despite their capacity to mediate FcγRII-dependent entry into B cells in vitro. *Vaccine* **25**, 729–740 (2007).
120. Pedersen, N. C. An update on feline infectious peritonitis: virology and immunopathogenesis. *Vet. J*. **201**, 123–132 (2014).
121. Rockx, B. et al. Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012–1015 (2020).
122. Munster, V. et al. Respiratory disease and virus shedding in rhesus macaques inoculated with SARS-CoV-2. Preprint at https://www.biorxiv.org/content/10.1101/2020.03.21.001628v1 (2020).
123. Chandrashekar, A. et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* https://doi.org/10.1126/science.abc4776 (2020).
Infection of rhesus macaques with SARS-CoV-2 and a comprehensive analysis of antibody neutralizing and Fc-mediated effector function showing multi-factorial correlation with protection against re-challenge.
124. Zhou, J. et al. Immunogenicity, safety, and protective efficacy of an inactivated SARS-associated coronavirus vaccine in rhesus monkeys. *Vaccine* **23**, 3202–3209 (2005).
125. Wang, Q. et al. Immunodominant SARS coronavirus epitopes in humans elicited both enhancing and neutralizing effects on infection in non-human primates. ACS *Infect. Dis*. **2**, 361–376 (2016).
126. Chen, Z. et al. Recombinant modified vaccinia virus Ankara expressing the spike glycoprotein of severe acute respiratory syndrome coronavirus induces protective neutralizing antibodies primarily targeting the receptor binding region. *J. Virol*. **79**, 2678–2688 (2005).
127. Qin, E. et al. Immunogenicity and protective efficacy in monkeys of purified inactivated Vero-cell SARS vaccine. *Vaccine* **24**, 1028–1034 (2006).
128. Clay, C. et al. Primary severe acute respiratory syndrome coronavirus infection limits replication but not lung inflammation upon homologous rechallenge. *J. Virol*. **86**, 4234–4244 (2012).
129. Muthumani, K. et al. A synthetic consensus anti-spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci. Transl. Med*. **7**, 301ra132 (2015).
130. Lan, J. et al. Recombinant receptor binding domain protein induces partial protective immunity in rhesus macaques against Middle East respiratory syndrome coronavirus challenge. *EBioMedicine* **2**, 1438–1446 (2015).
131. van Doremalen, N. et al. A single dose of ChAdOx1 MERS provides protective immunity in rhesus macaques. *Sci. Adv*. **6**, eaba8399 (2020).
132. Gao, Q. et al. Rapid development of an inactivated vaccine candidate for SARS-CoV-2. *Science* **369**, 77–81 (2020).
Protection of rhesus macaques against SARS-CoV-2 challenge after immunization with purified inactivated SARS-CoV-2 virus without evidence of ADE of disease.
133. Yu, J. et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* https://doi.org/10.1126/science.abc6284 (2020).
Immunization of rhesus macaques with DNA vaccines expressing forms of the SARS-CoV-2 S protein resulted in reduced infection following challenge after administration of full-length S protein without evidence of ADE of disease.
134. Miyoshi-Akiyama, T. et al. Fully human monoclonal antibody directed to proteolytic cleavage site in severe acute respiratory syndrome (SARS) coronavirus S protein neutralizes the virus in a rhesus macaque SARS model. *J. Infect. Dis*. **203**, 1574–1581 (2011).
135. Johnson, R. F. et al. 3B11-N, a monoclonal antibody against MERS-CoV, reduces lung pathology in rhesus monkeys following intratracheal inoculation of MERS-CoV Jordan-n3/2012. *Virology* **490**, 49–58 (2016).
136. de Wit, E. et al. Prophylactic and therapeutic efficacy of mAb treatment against MERS-CoV in common marmosets. *Antiviral Res*. **156**, 64–71 (2018).
137. de Wit, E. et al. Prophylactic efficacy of a human monoclonal antibody against MERS-CoV in the common marmoset. *Antiviral Res*. **163**, 70–74 (2019).
138. Chen, Z. et al. Human neutralizing monoclonal antibody inhibition of Middle East respiratory syndrome coronavirus replication in the common marmoset. *J. Infect. Dis*. **215**, 1807–1815 (2017).
139. Lam, J. H. et al. Dengue vaccine-induced $CD8^+$ T cell immunity confers protection in the context of enhancing, interfering maternal antibodies. *JCI Insight* **2**, e94500 (2017).

Acknowledgements The authors thank D. Ma for her contributions to preparing this review.

Author contributions A.M.A. and H.W.V. drafted the manuscript, K.F., M.A.S., A.C., R.S., C.H.-D., A.L. and D.C. edited the draft, and A.M.A. and H.W.V. prepared the final manuscript, which was reviewed and approved by all authors.

Competing interests The authors of this manuscript are employees of, or have other affiliations with, Vir Biotechnology. Vir had to choose how to proceed with mAbs to treat or prevent COVID-19 disease in the light of the evidence surrounding the possibility of ADE as detailed in this review. This review reflects the result of the team of authors carefully reviewing the literature to assess these choices and is provided as a service to the community. Vir has elected to test human mAbs with Fc activity preserved or enhanced, based on the lack of consistent evidence for ADE of disease noted above and evidence that the protective activities and potency of antibodies often involves antibody effector functions. We could have elected to take forward mAbs engineered to lack effector function, and so our antibody-related clinical programmes for SARS-CoV-2 could have moved forward regardless of the outcome of our review. A.M.A.'s contributions were part of her personal outside consulting arrangement with Vir Biotechnology and were not associated with Stanford University.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2538-8.
Correspondence and requests for materials should be addressed to A.M.A. or H.W.V.
Peer review information *Nature* thanks James Crowe, Gary Nabel and Stanley Perlman for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cold gas in the Milky Way's nuclear wind

https://doi.org/10.1038/s41586-020-2595-z

Received: 24 February 2020

Accepted: 8 June 2020

Enrico M. Di Teodoro[1,2,3] ✉, N. M. McClure-Griffiths[2], Felix J. Lockman[4] & Lucia Armillotta[2,5]

The centre of the Milky Way hosts several high-energy processes that have strongly affected the inner regions of our Galaxy. Activity from the super-massive black hole at the Galactic Centre, which is coincident with the radio source Sagittarius A*, and stellar feedback from the inner molecular ring[1] expel matter and energy from the disk in the form of a galactic wind[2]. Multiphase gas has been observed within this outflow, including hot highly ionized[3,4] (temperatures of about 10^6 kelvin), warm ionized[5,6] (10^4 to 10^5 kelvin) and cool atomic[7,8] (10^3 to 10^4 kelvin) gas. However, so far there has been no evidence of the cold dense molecular phase (10 to 100 kelvin). Here we report observations of molecular gas outflowing from the centre of our Galaxy. This cold material is associated with atomic hydrogen clouds travelling in the nuclear wind[8]. The morphology and the kinematics of the molecular gas, resolved on a scale of about one parsec, indicate that these clouds are mixing with the warmer medium and are possibly being disrupted. The data also suggest that the mass of the molecular gas outflow is not negligible and could affect the rate of star formation in the central regions of the Galaxy. The presence of this cold, dense and high-velocity gas is puzzling, because neither Sagittarius A* at its current level of activity nor star formation in the inner Galaxy seems to be a viable source for this material.

At a distance of only 8.2 kpc from the Sun (ref. [9]), the Galactic Centre provides a unique laboratory for studying the complex physical processes that occur within a galactic outflow. The 'Fermi bubbles'[2,10], two giant lobes extending up to ~10 kpc from the Galactic plane, are thought to outline the current boundaries of the Milky Way's nuclear wind. Several hundred neutral gas clouds have been found recently within this volume through observations of the atomic hydrogen (H I) line at a wavelength of $\lambda = 21$ cm (refs. [7,8]). Figure 1 shows a column density map of H I clouds in the nuclear wind[8] detected with the Green Bank Telescope (GBT). Although the bulk of the cloud population lies within the boundaries of the Fermi bubbles (green dashed line[11]), it has not been established whether this outflowing H I gas arises from the same event that generated the Fermi bubbles. These clouds were identified through their anomalous line-of-sight velocities, which are incompatible with Galactic rotation and can instead be described using a biconical wind model in which clouds accelerate from the Galactic Centre, reaching a maximum velocity of 330 km s^{-1} after about 2.5 kpc (refs. [8,12]). To assess whether outflowing H I structures carry molecular gas, we targeted two objects (hereafter, MW-C1 and MW-C2), highlighted by red boxes in Fig. 1, in the ^{12}CO(2 → 1) emission line at 230.538 GHz with the 12-m Atacama Pathfinder Experiment (APEX) telescope. These two clouds have relatively high H I column densities ($>10^{19}$ cm^{-2}) and show an elongated head-to-tail morphology along the direction pointing away from the Galactic Centre. We mapped both clouds in ^{12}CO(2 → 1) emission over a 15′ × 15′ field centred on the peak of the H I emission, at a spatial resolution of 28″ (full-width at half-maximum, FWHM), corresponding to ~1 pc at the distance of the Galactic Centre, and a spectral resolution of 0.25 km s^{-1}. These data revealed molecular gas outflowing from the centre of our Galaxy.

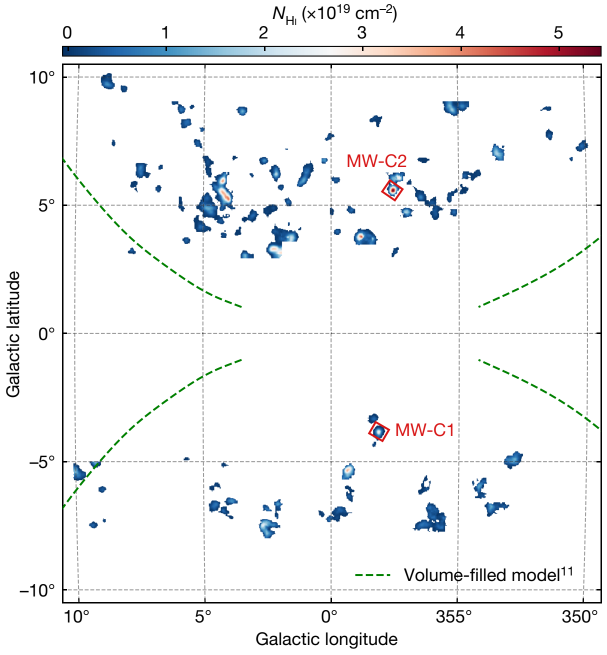


Fig. 1 | Atomic hydrogen gas outflowing from the Galactic Centre. The colour scale shows the column density of anomalous H I clouds in the Milky Way's nuclear wind, detected with GBT[8]. The green dashed line is the boundary of a volume-filled model for the Fermi bubbles[11]. The two H I clouds observed in the ^{12}CO(2 → 1) line with APEX are marked by red boxes.

---

[1]Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA. [2]Research School of Astronomy and Astrophysics, The Australian National University, Canberra, Australian Capital Territory, Australia. [3]Space Telescope Science Institute, Baltimore, MD, USA. [4]Green Bank Observatory, Green Bank, WV, USA. [5]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. ✉e-mail: editeodoro@jhu.edu

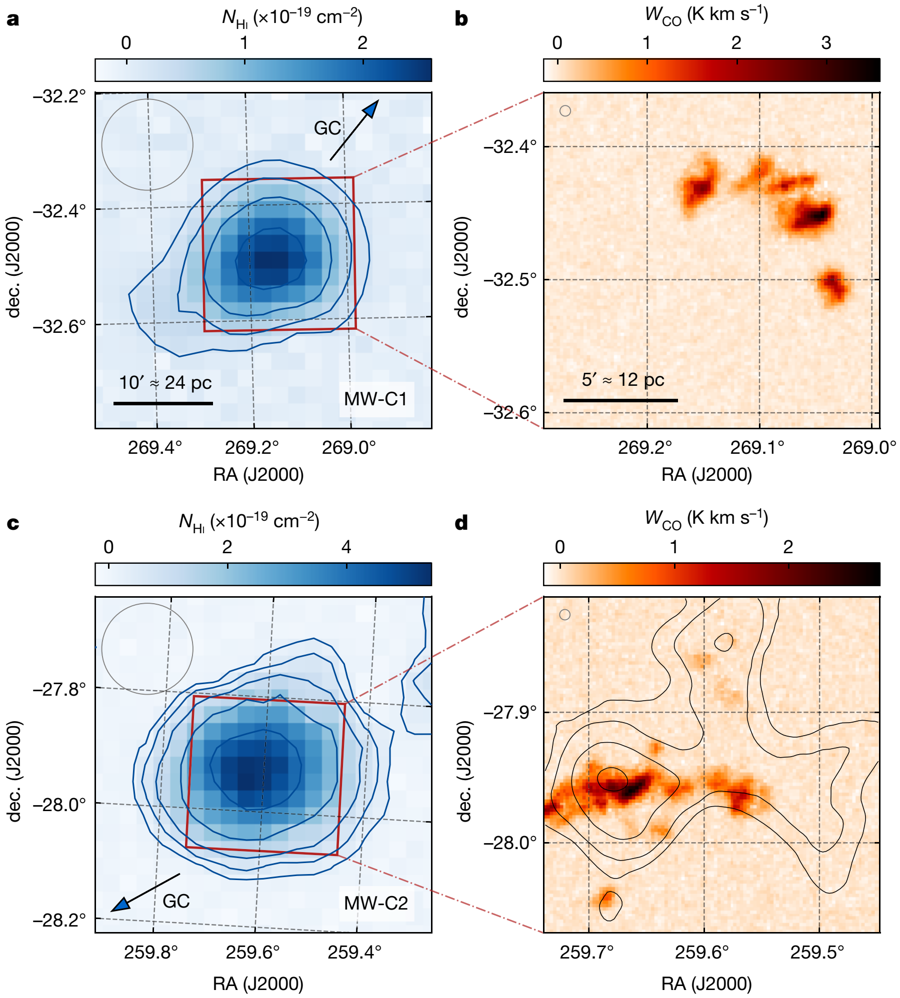


Fig. 2 | Atomic hydrogen and molecular gas in two clouds in the Milky Way's nuclear wind. a, **c**, H I column density maps from GBT data[8] at an angular resolution of 570″ for MW-C1 (**a**) and MW-C2 (**c**). Black arrows point towards the Galactic Centre (GC). Red boxes highlight the 15′ × 15′ fields observed with APEX. Contour levels are at $(0.2, 0.5, 1, 2, 4) \times 10^{19}$ cm^{-2}. **b**, **d**, ^{12}CO(2→1) integrated brightness temperature maps from APEX data at 28″ resolution for MW-C1 (**b**) and MW-C2 (**d**). H I contours at $(4, 8, 16, 24) \times 10^{20}$ cm^{-2} from ATCA data at 137″ resolution are overlaid on the MW-C2 map in **d**. The circles at the top left of each panel show the angular resolution of the telescopes. RA, right ascension; dec., declination; W_{CO}, integrated brightness temperature.

Figure 2 shows H I column density maps (Fig. 2a, c) from GBT observations and integrated brightness temperature maps (Fig. 2b, d) from the ^{12}CO(2→1) line obtained with APEX for MW-C1 and MW-C2. Higher-resolution H I data from the Australia Telescope Compact Array (ATCA) for MW-C2 are also overlaid as contours on the CO map. CO velocity fields and three representative spectra across each field are presented in Fig. 3. CO emission is detected in both H I clouds, with substantial morphological and kinematical differences between them. MW-C1 shows five distinct compact clumps of molecular gas concentrated towards the part of the H I cloud that faces the Galactic Centre (arrows in Fig. 2). At least three clumps have a velocity gradient along the direction pointing towards the tail of the H I cloud. All the CO emission in MW-C1 lies in the local-standard-of-rest (LSR) velocity range $V_{LSR} \approx 160$–170 km s^{-1}. Typical FWHM line widths are ~2–3 km s^{-1} (see spectra in Fig. 3). By contrast, in MW-C2 most of the CO emission is distributed along a filament-like structure, with some fainter and more diffuse clumps in the region away from the Galactic Centre. CO emission is spread over a larger velocity range than in MW-C1, spanning 30 km s^{-1} over $V_{LSR} \approx 250$–280 km s^{-1}, and the velocity field does not show any clear ordered motion. ^{12}CO(2→1) line profiles in MW-C2 are much broader than in MW-C1, with an FWHM ranging from ~5 km s^{-1} to 12 km s^{-1}.

The observed features indicate that cold gas in MW-C2 is interacting and mixing with the surrounding medium more efficiently than in MW-C1, resulting in a more turbulent molecular gas. An interpretation of the differences in the morphokinematics of the molecular gas in the two clouds is that we are witnessing two evolutionary stages of a cold cloud being disrupted by interaction with a hot flow. Our idealized biconical wind model[12] with a maximum wind velocity of 330 km s^{-1} places MW-C1 at a distance of 0.8 kpc and MW-C2 at a distance of 1.8 kpc from the Galactic Centre, implying that MW-C2 may have been within the nuclear outflow twice as long (7 Myr, versus 3 Myr for MW-C1). Our model also predicts that MW-C2 is moving faster than MW-C1 (~300 km s^{-1} versus ~240 km s^{-1}). In the classical picture, in which cold gas is entrained in the hot wind, MW-C1 may therefore represent an early stage of the interaction with the surrounding medium, at which molecular gas is still relatively intact and undisturbed near the initial dense core; whereas molecular gas in MW-C2 could have been stripped off from its core, resulting in a disordered morphology/velocity field and broader linewidths. However, the observed characteristics of the two clouds may also be explained in terms of different local conditions of the hot outflow. A larger and more complete sample of molecular gas detections in outflowing clouds is needed to provide a more robust picture.

The two clouds analysed in this work have atomic gas masses of $M_{at} \approx 220 M_\odot$ (MW-C1) and $M_{at} \approx 800 M_\odot$ (MW-C2), as derived from the GBT H I data ($M_\odot$, mass of the Sun). All mass measurements from

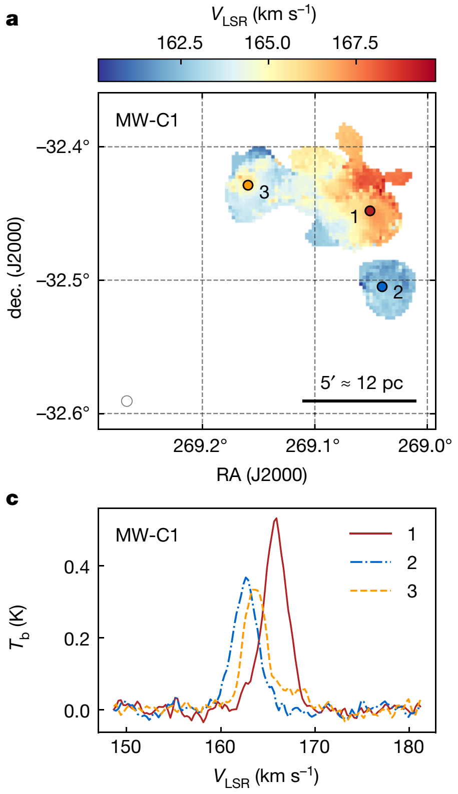


Fig. 3 | Molecular gas kinematics in MW-C1 and MW-C2. a, **b**, Velocity fields derived from a Gaussian fit to the ^{12}CO(2→1) data for MW-C1 (**a**) and MW-C2 (**b**). **c**, **d**, ^{12}CO(2→1) spectra for MW-C1 (**c**) and MW-C2 (**d**) at the positions labelled in **a** and **b**. We note the differences in velocity spread and line shape between MW-C1 and MW-C2. T_b is the line brightness temperature.

observations are scaled by a factor of 1.36 to account for the presence of helium. It is not straightforward to estimate the mass of molecular matter, because the gas may have considerable opacity in the ^{12}CO(2→1) line and the appropriate CO-to-H$_2$ conversion factor X_{CO} in the Milky Way's wind is unknown. We used the observed CO integrated brightness temperatures, cloud radii and line widths to constrain the acceptable X_{CO} values by means of chemical and thermal modelling of a cloud undergoing dissociation by photons and cosmic rays. We found that X_{CO} for the ^{12}CO(2→1) transition in our clouds lies in the range (~2–40) × 10^{20} cm^{-2} (K km s^{-1})$^{-1}$. The lowest value, $X_{CO} = 2 \times 10^{20}$ cm^{-2} (K km s^{-1})$^{-1}$, is consistent with the Galactic conversion factor[13], and was used to derive lower limits to the molecular gas mass M_{mol}. We obtained $M_{mol} \geq 380 M_{\odot}$ for MW-C1 and $M_{mol} \geq 375 M_{\odot}$ for MW-C2, implying molecular-to-total gas mass fractions of $f_{mol} = M_{mol}/(M_{mol} + M_{at}) \geq 0.64$ and $f_{mol} \geq 0.32$, respectively. We emphasize that these values are lower bounds and the molecular gas mass may be higher by a factor of ten. As a consequence, the total mass of molecular gas in the nuclear wind of the Milky Way is large. Under the conservative assumption of an average $f_{mol} \approx 0.3$–0.5 for all outflowing H I clouds in the GBT sample, and using an atomic outflow rate[8] of $\dot{M}_{at} \approx 0.1 M_{\odot}$ yr^{-1}, we estimated an outflow rate of $\dot{M}_{at} \geq (0.05$–$0.1) M_{\odot}$ yr^{-1} in molecular gas. This value is of the same order of magnitude as the star formation rate (SFR) of the Central Molecular Zone[14] (CMZ), implying a molecular gas loading factor $\eta = \dot{M}_{mol}$/SFR at least of the order of unity at a distance of ~1 kpc from the Galactic plane, similar to that estimated in nearby starburst galaxies[15]. This cold outflow affects the gas cycle in the inner Galaxy and may constitute an important mechanism that regulates the star formation activity in the CMZ.

From a theoretical point of view, such a large amount of high-velocity molecular gas is puzzling[16]. It is believed that cool gas in a disk can be lifted and accelerated by both drag force from a hot outflow[17] and by radiation pressure[18]. This requires a source of strong thermal feedback and/or radiation feedback. The Milky Way does not currently have an active galactic nucleus (AGN), nor is the SFR of the inner Galaxy comparable to that of starburst galaxies with known molecular winds (for example, NGC253)[15]. Current simulations of AGN-driven winds have focused on very powerful AGNs[19,20] and there have been no investigations studying whether a relatively small black hole like Sagittarius A* could expel large amounts of cold gas, even if it had undergone a period of activity in the recent past. On the other hand, the current SFR of the CMZ is not large enough to explain the estimated outflow rate of cold gas[21], and no observational evidence so far suggests a sizable change in the SFR of the CMZ in the last few million years[22]. A scenario in which the star formation in the CMZ is episodic on a longer cycle[23,24] (10–50 Myr) and is currently near a minimum might help to partly reconcile the observed and predicted cool gas mass loading rates, although our wind model suggests that the lifetimes of cold clouds are shorter than 10 Myr. Cosmic rays are also believed to contribute to the pressure on cold gas[25], but their role is only just starting to be understood and needs observational constraints. Moreover, in either an AGN- or a starburst-driven wind, the extent to which cold gas survives under acceleration is a matter of debate[17,26], and several different mechanisms have been investigated to extend the lifetime of cool gas in a hot wind (for example, magnetic fields[27] and thermal conduction[28]). An alternative scenario has been recently proposed in which high-velocity cool neutral gas (temperature $T < 10^4$ K) forms directly within the outflow as a consequence of mixing between slow-moving cool clouds and the fast-moving hot wind[29,30]. This mechanism overcomes the problem of accelerating dense material without disrupting it, and may explain the high velocities observed in cool outflows. However, current simulations cannot trace the gas down to the molecular phase.

In conclusion, this detection of ouflowing cold molecular gas in the Milky Way is a challenge for current theories of galactic winds in regular star-forming galaxies, because none of the above processes seems able to easily explain the presence of fast molecular gas in the Milky Way's wind. Targeted observations of molecular gas tracers in the Milky Way's nuclear wind are expected to contribute considerably to our understanding of these fascinating phenomena.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2595-z.

1. Molinari, S. et al. A 100 pc elliptical and twisted ring of cold and dense molecular clouds revealed by *Herschel* around the Galactic center. *Astrophys. J. Lett.* **735**, 33 (2011).
2. Bland-Hawthorn, J. & Cohen, M. The large-scale bipolar wind in the Galactic center. *Astrophys. J.* **582**, 246–256 (2003).
3. Kataoka, J. et al. *Suzaku* observations of the diffuse X-ray emission across the Fermi Bubbles' edges. *Astrophys. J.* **779**, 57 (2013).
4. Ponti, G. et al. An X-ray chimney extending hundreds of parsecs above and below the Galactic Centre. *Nature* **567**, 347–350 (2019).
5. Fox, A. J. et al. Probing the Fermi Bubbles in ultraviolet absorption: a spectroscopic signature of the Milky Way's biconical nuclear outflow. *Astrophys. J.* **799**, L7 (2015).
6. Bordoloi, R. et al. Mapping the nuclear outflow of the Milky Way: studying the kinematics and spatial extent of the northern Fermi Bubble. *Astrophys. J.* **834**, 191 (2017).
7. McClure-Griffiths, N. M. et al. Atomic hydrogen in a Galactic center outflow. *Astrophys. J. Lett*. **770**, 4 (2013); erratum **884**, 27 (2019).
8. Di Teodoro, E. M. et al. Blowing in the Milky Way wind: neutral hydrogen clouds tracing the Galactic nuclear outflow. *Astrophys. J.* **855**, 33 (2018).
9. Gravity Collaboration. Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic centre massive black hole. *Astron. Astrophys.* **636**, L5 (2020).
10. Su, M., Slatyer, T. R. & Finkbeiner, D. P. Giant gamma-ray bubbles from Fermi-LAT: active galactic nucleus activity or bipolar Galactic wind? *Astrophys. J.* **724**, 1044–1082 (2010).
11. Miller, M. J. & Bregman, J. N. The interaction of the Fermi Bubbles with the Milky Way's hot gas halo. *Astrophys. J.* **829**, 9 (2016).
12. Lockman, F. J., Di Teodoro, E. M. & McClure-Griffiths, N. M. Observation of acceleration of HI clouds within the Fermi Bubbles. *Astrophys. J.* **888**, 51 (2020).
13. Bolatto, A. D., Wolfire, M. & Leroy, A. K. The CO-to-H_2 conversion factor. *Annu. Rev. Astron. Astrophys*. **51**, 207–268 (2013).
14. Longmore, S. N. et al. Variations in the Galactic star formation rate and density thresholds for star formation. *Mon. Not. R. Astron. Soc*. **429**, 987–1000 (2013).
15. Bolatto, A. D. et al. Suppression of star formation in the galaxy NGC253 by a starburst-driven molecular wind. *Nature* **499**, 450–453 (2013).
16. Veilleux, S., Maiolino, R., Bolatto, A. D. & Aalto, S. Cool outflows in galaxies and their implications. *Annu. Rev. Astron. Astrophys*. **28**, 2 (2020).
17. Scannapieco, E. & Brüggen, M. The launching of cold clouds by Galaxy outflows. I. Hydrodynamic interactions with radiative cooling. *Astrophys. J.* **805**, 158 (2015).
18. Thompson, T. A., Fabian, A. C., Quataert, E. & Murray, N. Dynamics of dusty radiation-pressure-driven shells and clouds: fast outflows from galaxies, star clusters, massive stars, and AGN. *Mon. Not. R. Astron. Soc*. **449**, 147–161 (2015).
19. Mukherjee, D., Bicknell, G. V., Sutherland, R. & Wagner, A. Relativistic jet feedback in high-redshift galaxies – I. Dynamics. *Mon. Not. R. Astron. Soc*. **461**, 967–983 (2016).
20. Richings, A. J. & Faucher-Giguère, C.-A. Radiative cooling of swept-up gas in AGN-driven galactic winds and its implications for molecular outflows. *Mon. Not. R. Astron. Soc*. **478**, 3100–3119 (2018).
21. Armillotta, L., Krumholz, M. R., Di Teodoro, E. M. & McClure-Griffiths, N. M. The life cycle of the Central Molecular Zone – I. Inflow, star formation, and winds. *Mon. Not. R. Astron. Soc*. **490**, 4401–4418 (2019).
22. Barnes, A. T. et al. Star formation rates and efficiencies in the Galactic Centre. *Mon. Not. R. Astron. Soc*. **469**, 2263–2285 (2017).
23. Krumholz, M. R., Kruijssen, J. M. D. & Crocker, R. M. A dynamical model for gas flows, star formation and nuclear winds in galactic centres. *Mon. Not. R. Astron. Soc*. **466**, 1213–1233 (2017).
24. Armillotta, L., Krumholz, M. R. & Di Teodoro, E. M. The life cycle of the Central Molecular Zone – II. Distribution of atomic and molecular gas tracers. *Mon. Not. R. Astron. Soc*. **493**, 5273–5289 (2020).
25. Girichidis, P., Naab, T., Hanasz, M. & Walch, S. Cooler and smoother – the impact of cosmic rays on the phase structure of galactic outflows. *Mon. Not. R. Astron. Soc*. **479**, 3042–3067 (2018).
26. Zhang, D., Thompson, T. A., Quataert, E. & Murray, N. Entrainment in trouble: cool cloud acceleration and destruction in hot supernova-driven galactic winds. *Mon. Not. R. Astron. Soc*. **468**, 4801–4814 (2017).
27. McCourt, M., O'Leary, R. M., Madigan, A.-M. & Quataert, E. Magnetized gas clouds can survive acceleration by a hot wind. *Mon. Not. R. Astron. Soc*. **449**, 2–7 (2015).
28. Armillotta, L., Fraternali, F., Werk, J. K., Prochaska, J. X. & Marinacci, F. The survival of gas clouds in the circumgalactic medium of Milky Way-like galaxies. *Mon. Not. R. Astron. Soc*. **470**, 114–125 (2017).
29. Gronke, M. & Oh, S. P. The growth and entrainment of cold gas in a hot wind. *Mon. Not. R. Astron. Soc*. **480**, L111–L115 (2018).
30. Schneider, E. E., Ostriker, E. C., Robertson, B. E. & Thompson, T. A. The physical nature of starburst-driven Galactic outflows. *Astrophys. J.* **895**, 43 (2020).

Methods

Observations and data reduction

Observations of the ^{12}CO(2 → 1) emission line at 230.538 GHz were made with the 12-m APEX antenna[31] using the PI230 heterodyne receiver (ESO project 0104.B-0106A; principal investigator E.M.D.T.). The spectrometer[32] covers a bandwidth of 8 GHz at a spectral resolution of 61 kHz, corresponding to a velocity resolution of about 0.08 km s^{-1} at 230 GHz. The beam size at this frequency is 27.8″ (FWHM), the main-beam efficiency is 0.72 and the jansky-to-kelvin conversion factor is 40 ± 3. We observed our targets in on-the-fly position-switching mode, integrating for 1 s every 9″. Both fields were 15′ × 15′ wide, centred at (RA, dec.)$_{\rm J2000}$ = (17 h 56 min 34.0 s, −32° 29′ 14″) for MW-C1 and at (RA, dec.)$_{\rm J2000}$ = (17 h 18 min 22.2 s, −27° 56′ 28″) for MW-C2. The observed regions are shown in red boxes in Figs. 1, 2. The total integration time was approximately 25 h for each field. Throughout the observing session (September to November 2019), the precipitable water vapour varied between 0.6 mm and 3 mm.

We reduced the data using the Continuum and Line Analysis Single-dish Software (CLASS) from the GILDAS package[33]. A first-order baseline was subtracted from the calibrated spectra by interpolating the channels outside the velocity windows in which we expected to see the emission based on the H I observations. The spectra were then smoothed in velocity and mapped onto a grid with a pixel size of 9″ and a channel width of 0.25 km s^{-1}. The root-mean-square noise ($\sigma_{\rm rms}$) in the final data cubes was 65 mK and 55 mK for MW-C1 and MW-C2, respectively, in a 0.25 km s^{-1} channel.

Atomic gas and molecular gas mass

The H I GBT data[8] and the ^{12}CO(2 → 1) APEX data were analysed to estimate the atomic and molecular gas masses, respectively. First, the three-dimensional source finder DUCHAMP[34] was applied to the data cubes to identify regions of sizable emission. During this process, we set a primary threshold to identify emission peaks at $5\sigma_{\rm rms}$ and reconstructed sources by adding pixels down to a secondary threshold of $2.5\sigma_{\rm rms}$.

The column density at a given position (x, y) on the sky can be written as:

$$N_{\rm H}(x,y) = C\int T_{\rm b}(x,y,v)\,{\rm d}v, \qquad (1)$$

where the integral considers pixels in only one detection, $T_{\rm b}$ is the line brightness temperature, dv is the channel width (1 km s^{-1} for GBT and 0.25 km s^{-1} for APEX) and C is a constant. For the H I line, under the assumption that the gas is optically thin, the constant is[35] $C = 1.82 \times 10^{18}$ cm^{-2} (K km s^{-1})$^{-1}$. For CO lines, this constant is also known as the CO-to-H$_2$ conversion factor $X_{\rm CO}$ (ref. [13]). Because the conversion factor in the nuclear wind cannot be constrained with existing data, we used the value estimated in molecular clouds in the disk of the Milky Way[36], $X_{\rm CO} = 2 \times 10^{20}$ cm^{-2} (K km s^{-1})$^{-1}$. We checked this $X_{\rm CO}$ value against the predictions of radiative-transfer models, described in the next section, and found that the Galactic value is probably a lower limit for clouds in the nuclear wind.

The total mass of gas can be calculated as:

$$M = 1.36 m D^2 \int N_{\rm H}(x,y)\,{\rm d}x{\rm d}y, \qquad (2)$$

where the factor 1.36 takes into account helium, $D \approx 8.2$ kpc is the adopted distance to the clouds, m is the mass of atomic/molecular hydrogen for atomic/molecular gas, dx and dy are the pixel sizes in radians (105″ for GBT, 9″ for APEX). The observed properties and estimated masses are summarized in the Extended Data Table 1.

Radiative-transfer models

We used the chemistry and radiative-transfer code DESPOTIC[37] to constrain the CO-to-H$_2$ conversion factor of the clouds. DESPOTIC computes the chemical and thermal state of an optically thick cloud given its volume density and column density. The turbulent velocity dispersion of the gas was assumed to be 1–5 km s^{-1} (see Fig. 3) in our modelling. The chemical equilibrium calculation uses solar abundances for dust and all elements in the H–C–O chemical network[38], whereas the thermal equilibrium calculation includes heating by cosmic rays, the grain photoelectric effect, cooling by the H I, C I, C II, O I and CO lines, and collisional energy exchange between dust and gas. Level populations were calculated using an escape probability method, with escape probabilities estimated using the spherical geometry option of DESPOTIC.

We investigated different combinations of the interstellar radiation field χ and the cosmic-ray ionization rate ζ through a set of DESPOTIC models using $\log(\chi/G_0) = [-1, 0, 1, 2]$, where G_0 is the solar radiation field[39] and $\log[\zeta\,({\rm s}^{-1})] = [-16, -15, -14]$. The interstellar radiation field was varied between subsolar ($\chi \approx 0.1G_0$) and highly supersolar ($\chi \approx 100G_0$) values, representative of a highly star-forming environment like the CMZ. The cosmic-ray ionization rate ranges from the value measured in the solar neighbourhood[40] ($\zeta \approx 10^{-16}$ s^{-1}) to the estimated upper limit for the CMZ[41] ($\zeta \approx 10^{-14}$ s^{-1}). We stress that our CO clouds lie at about 1 kpc from the Galactic plane and that both the interstellar radiation field and the cosmic-ray ionization rate are expected to drop with distance from the disk. Therefore, although the estimated values of χ and ζ in the CMZ are orders of magnitude higher than in the solar neighbourhood, models with intermediate interstellar radiation fields and cosmic-ray ionization rates should be more representative of the conditions high in the Milky Way's wind.

For each model, DESPOTIC returned the ^{12}CO(2 → 1) integrated brightness temperatures ($W_{\rm CO}$) as a function of the number density ($n_{\rm H2}$) and column density ($N_{\rm H2}$) of molecular hydrogen. We only considered solutions consistent with the observed integrated brightness temperature (1–5 K km s^{-1}; see Fig. 2) and observed cloud radius of $R = 0.75n_{\rm H2}/N_{\rm H2} = 1$–5 pc, and we calculated the expected CO-to-H$_2$ conversion factor $X_{\rm CO} = N_{\rm H2}/W_{\rm CO}$ for the ^{12}CO(2 → 1) transition. We found that there are no acceptable solutions for a strong interstellar radiation field ($\log(\chi/G_0) \geq 1$), which indicates that molecular clouds with the observed properties cannot exist in the presence of a CMZ-like radiation field. Instead, models with solar and subsolar radiation fields returned solutions compatible with the observational constraints for any cosmic-ray ionization field. An interstellar radiation field weaker than the one produced in the CMZ is therefore more representative of the environment at 1 kpc above the Galactic Centre. The predicted $X_{\rm CO}$ varies by an order of magnitude, ranging between ~2×10^{20} cm^{-2} (K km s^{-1})$^{-1}$ and ~4×10^{21} cm^{-2} (K km s^{-1})$^{-1}$, depending on the combination of radiation field and cosmic-ray ionization rate. The value of $X_{\rm CO} = 2\times10^{20}$ cm^{-2} (K km s^{-1})$^{-1}$ that is commonly assumed in the Milky Way disk[13] and used in this study is consistent with the smallest values returned by our radiative-transfer models, obtained with a weak, subsolar radiation field and a solar-like cosmic-ray ionization rate of $\zeta = 10^{-16}$ s^{-1}. As a consequence, the molecular gas masses calculated in this work probably represent lower limits to the real cold gas mass in our CO clouds.

Wind kinematic model

To estimate the position, velocity and lifetime of MW-C1 and MW-C2, we used a biconical wind model[8,12] calibrated on the full population of H I clouds. This model is based on the assumption that clouds were launched from a small region close to the centre of the Galaxy and are moving with a purely radial velocity $V_{\rm w}(r)$, where r is the distance from the Galactic Centre. For simplicity, we considered models of the form:

$$V_{\rm w}(r) = \begin{cases} V_{\rm i} + (V_{\rm max} - V_{\rm i})\dfrac{r}{r_{\rm s}} & \text{for } r < r_{\rm s} \\ V_{\rm max} & \text{for } r \geq r_{\rm s} \end{cases}, \qquad (3)$$

where V_i is the initial velocity at $r = 0$ and r_s is the scale distance at which the maximum velocity V_{max} is reached. Equation (3) describes a kinematic model in which clouds are subjected to a constant acceleration up to r_s and maintain a constant velocity at distances $r \geq r_s$. Although equation (3) is purely empirical and chosen to reproduce the H I data, recent hydrodynamical simulations of starburst-driven winds have found qualitatively similar trends for the velocity of the cool gas with distance[30]. The LSR velocity V_{LSR} of a cloud travelling in the wind and seen at Galactic coordinates (l, b) can be written as:

$$V_{LSR}(l, b, r) = V_w(r)[\sin\varphi \sin b - \cos\varphi \cos b \cos(l + \theta)] - V_0 \sin l \sin b,$$

where the polar angle φ and the azimuthal angle θ can be easily written as a function of (l, b, r) (ref. [8]) and $V_0 = 240$ km s^{-1} is the rotation velocity of the LSR around the Galactic Centre[42]. In our model, clouds are restricted inside a bicone with half-opening angle φ_{max}. We constrained the four free parameters of this model, V_i, V_{max}, r_s and φ_{max}, by matching the LSR velocity distributions predicted by our model with that observed from the H I cloud population[8,12]. Our fiducial model is a biconical wind with opening angle $\varphi_{max} = 70°$, where clouds accelerate from an initial velocity of $V_i = 200$ km s^{-1} to a maximum velocity of $V_{max} = 330$ km s^{-1} at $r_s = 2.5$ kpc. According to this wind model, MW-C1 and MW-C2 have travelled a distance of 0.8 kpc and 1.8 kpc from the Galactic Centre in about 3 Myr and 7 Myr, and their current outflow velocity is about 240 km s^{-1} and 300 km s^{-1}, respectively.

Data availability

The APEX raw datasets analysed for this study will be available at the end of the proprietary period (September 2020) on the ESO archive, http://archive.eso.org/eso/eso archive main.html. The GBT raw datasets are publicly available at the NRAO archive, https://science.nrao.edu/facilities/gbt/software-and-tools. Fully reduced data are available from the corresponding author on reasonable request.

Code availability

The software used in this work is publicly available. The GILDAS/CLASS packages for submillimetre data reduction can be found at https://www.iram.fr/IRAMFR/GILDAS. The DUCHAMP source finder can be downloaded from https://www.atnf.csiro.au/people/Matthew.Whiting/Duchamp. The DESPOTIC radiative-transfer code is available at https://bitbucket.org/krumholz/despotic.

31. Güsten, R. et al. The Atacama Pathfinder EXperiment (APEX) – a new submillimeter facility for southern skies. *Astron. Astrophys*. **454**, L13–L16 (2006).
32. Klein, B. et al. High-resolution wide-band fast Fourier transform spectrometers. *Astron. Astrophys*. **542**, L3 (2012).
33. Gildas Team. GILDAS: Grenoble Image and Line Data Analysis Software. *Astrophysics Source Code Library* http://www.ascl.net/1305.010 (2013).
34. Whiting, M. T. DUCHAMP: a 3D source finder for spectral-line data. *Mon. Not. R. Astron. Soc*. **421**, 3242–3256 (2012).
35. Roberts, M. S. Radio observations of neutral hydrogen in galaxies. In *Galaxies and the Universe* (eds Sandage, A., Sandage, M. & Kristian, J.) 309–358 (Univ. of Chicago Press, 1975).
36. Heyer, M., Krawczyk, C., Duval, J. & Jackson, J. M. Re-examining Larson's scaling relationships in galactic molecular clouds. *Astrophys. J*. **699**, 1092–1103 (2009).
37. Krumholz, M. R. DESPOTIC – a new software library to Derive the Energetics and SPectra of Optically Thick Interstellar Clouds. *Mon. Not. R. Astron. Soc*. **437**, 1662–1680 (2014).
38. Gong, M., Ostriker, E. C. & Wolfire, M. G. A simple and accurate network for hydrogen and carbon chemistry in the interstellar medium. *Astrophys. J*. **843**, 38 (2017); erratum **866**, 163 (2018).
39. Draine, B. T. Photoelectric heating of interstellar gas. *Astrophys. J. Suppl. Ser*. **36**, 595–619 (1978).
40. Indriolo, N. & McCall, B. J. Investigating the cosmic-ray ionization rate in the galactic diffuse interstellar medium through observations of H_3^+. *Astrophys. J*. **745**, 91 (2012).
41. Oka, T. et al. The central 300 pc of the Galaxy probed by infrared spectra of H_3^+ and CO. I. Predominance of warm and diffuse gas and high H_2 ionization rate. *Astrophys. J*. **883**, 54 (2019).
42. Bland-Hawthorn, J. & Gerhard, O. The Galaxy in context: structural, kinematic, and integrated properties. *Annu. Rev. Astron. Astrophys*. **54**, 529–596 (2016).

Acknowledgements E.M.D.T. and L.A. thank E. Ostriker, C.-G. Kim and J.-G. Kim for discussions and M. Krumholz for support with the DESPOTIC code. E.M.D.T. was supported by the US National Science Foundation under grant 1616177. E.M.D.T. and N.M.M.-G. acknowledge the support of the Australian Research Council (ARC) through grant DP160100723. N.M.M.-G. acknowledges funding from the ARC via Future Fellowship FT150100024. CO observations were made with APEX under ESO proposal 0104.B-0106A. APEX is a collaboration between Max-Planck-Institut für Radioastronomie, the European Southern Observatory and the Onsala Space Observatory. The Green Bank Observatory is a facility of the US National Science Foundation operated under a cooperative agreement by Associated Universities, Inc. The ATCA is part of the Australia Telescope National Facility, which is funded by the Australian Government for operation as a National Facility managed by CSIRO.

Author contributions E.M.D.T., N.M.M.-G. and F.J.L. developed the idea for the project. E.M.D.T. reduced and analysed the APEX data, L.A. ran the radiative-transfer models. E.M.D.T. wrote the paper with direct contributions from N.M.M.-G., F.J.L. and L.A. All authors reviewed the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2595-z.
Correspondence and requests for materials should be addressed to E.M.D.T.
Peer review information *Nature* thanks Mark Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.

Extended Data Table 1 | Properties of molecular gas clouds outflowing from the Galactic Centre

	ℓ ($^\circ$)	b ($^\circ$)	z (kpc)	r (kpc)	$T_{\rm b,peak}$ (K)	FWHM ($\rm km\,s^{-1}$)	Vel. range ($\rm km\,s^{-1}$)	$M_{\rm mol}$ ($\rm M_\odot$)	$M_{\rm at}$ ($\rm M_\odot$)
MW-C1	358.14	-3.84	0.6	0.8	1.5	$2-3$	$160-170$	380	220
MW-C2	357.58	5.56	0.9	1.8	0.5	$5-12$	$250-280$	375	800

Shown are the Galactic coordinates (*l*, *b*); the height from the Galactic plane (*z*); the distance from the Galactic Centre (*r*) from our biconical outflow model[12]; the peak ^{12}CO(2 → 1) brightness temperature ($T_{\rm b,peak}$); typical CO line widths (FWHM); the velocity range of the CO line in the LSR; the lower limits to the molecular masses ($M_{\rm mol}$), derived from ^{12}CO(2 → 1) data; and atomic gas masses ($M_{\rm at}$), derived from H I data. Masses include helium.

Quantum error correction of a qubit encoded in grid states of an oscillator

https://doi.org/10.1038/s41586-020-2603-3

Received: 8 November 2019

Accepted: 12 June 2020

P. Campagne-Ibarcq[1,3,6]✉, A. Eickbusch[1,6], S. Touzard[1,4,6]✉, E. Zalys-Geller[1], N. E. Frattini[1], V. V. Sivak[1], P. Reinhold[1], S. Puri[1], S. Shankar[1,5], R. J. Schoelkopf[1], L. Frunzio[1], M. Mirrahimi[2] & M. H. Devoret[1]✉

The accuracy of logical operations on quantum bits (qubits) must be improved for quantum computers to outperform classical ones in useful tasks. One method to achieve this is quantum error correction (QEC), which prevents noise in the underlying system from causing logical errors. This approach derives from the reasonable assumption that noise is local, that is, it does not act in a coordinated way on different parts of the physical system. Therefore, if a logical qubit is encoded non-locally, we can—for a limited time—detect and correct noise-induced evolution before it corrupts the encoded information[1]. In 2001, Gottesman, Kitaev and Preskill (GKP) proposed a hardware-efficient instance of such a non-local qubit: a superposition of position eigenstates that forms grid states of a single oscillator[2]. However, the implementation of measurements that reveal this noise-induced evolution of the oscillator while preserving the encoded information[3–7] has proved to be experimentally challenging, and the only realization reported so far relied on post-selection[8,9], which is incompatible with QEC. Here we experimentally prepare square and hexagonal GKP code states through a feedback protocol that incorporates non-destructive measurements that are implemented with a superconducting microwave cavity having the role of the oscillator. We demonstrate QEC of an encoded qubit with suppression of all logical errors, in quantitative agreement with a theoretical estimate based on the measured imperfections of the experiment. Our protocol is applicable to other continuous-variable systems and, in contrast to previous implementations of QEC[10–14], can mitigate all logical errors generated by a wide variety of noise processes and facilitate fault-tolerant quantum computation.

The qubit encoding proposed by GKP is based on grid patterns in phase space, which only emerge by interfering periodically spaced position eigenstates with adequate phase relationships, as shown in Fig. 1. The resulting 'grid-state' code belongs to the class of stabilizer codes. In the stabilizer formalism of QEC, the measurement of chosen operators—the stabilizers—reveals unambiguously the action of undesired noise without disturbing the state of the logical qubit. As a consequence of this latter condition, the stabilizers must commute with all observables of the logical qubit, which are combinations of the logical Pauli operators. For the grid-state code, these operators are phase-space displacements, defined as $\mathbf{D}(\beta) = \exp(-\mathrm{i}\mathrm{Re}(\beta)\mathbf{p} + \mathrm{i}\mathrm{Im}(\beta)\mathbf{q})$, where $\mathbf{q}$ and $\mathbf{p}$ are the conjugated position and momentum operators, such that $[\mathbf{q}, \mathbf{p}] = \mathrm{i}$. For example, the stabilizers of the canonical square grid-state code are $\mathbf{S}_a = \mathbf{D}(a = 2\sqrt{\pi})$ and $\mathbf{S}_b = \mathbf{D}(b = 2\mathrm{i}\sqrt{\pi})$, and the Pauli operators are $\mathbf{X} = \mathbf{D}(a/2)$, $\mathbf{Z} = \mathbf{D}(b/2)$ and $\mathbf{Y} = \mathbf{D}((a+b)/2)$. The phase of the stabilizers encodes no information about the logical qubit, but reveals the momentum shifts modulo $2\pi/|a|$ and the position shifts modulo $2\pi/|b|$. Thus, shifts that are smaller than a quarter of a grid period are unambiguously identified and can be corrected. Because usual decoherence processes, such as photon relaxation[15,16], pure dephasing and spurious nonlinearities, result in a continuous evolution of the quasi-probability distribution in phase space[2,17], shifts of order a, b do not occur instantaneously. Therefore, if the stabilizers are measured frequently enough, noise-induced shifts can be detected and corrected, which inhibits all logical errors.

However, in contrast to this description, which is based on ideal position eigenstates, physically realizable code states do not extend infinitely in phase space; they are superpositions of periodically spaced squeezed states of width σ, with a Gaussian overall envelope of width $\Delta = 1/(2\sigma)$ (see Fig. 1a). These states are still approximate eigenstates of the stabilizers, such that $|\langle \mathbf{S}_{a,b} \rangle| \approx 1$. Any pair of orthogonal logical states are shifted from one to the other in phase space (for example, by $a/2$ for $|\pm Z_L\rangle$ and $b/2$ for $|\pm X_L\rangle$). For sufficient squeezing, their supports do not considerably overlap, the logical qubit is well defined, and a QEC protocol can be directly adapted from the ideal case.

---

[1] Department of Applied Physics, Yale University, New Haven, CT, USA. [2] Quantic Team, INRIA Paris, Paris, France. [3] Present address: Quantic Team, INRIA Paris, Paris, France. [4] Present address: Centre for Quantum Technologies, National University of Singapore, Singapore, Singapore. [5] Present address: Department of Electrical and Computer Engineering, University of Texas, Austin, TX, USA. [6] These authors contributed equally: P. Campagne-Ibarcq, A. Eickbusch, S. Touzard. ✉e-mail: philippe.campagne-ibarcq@inria.fr; steven.touzard@nus.edu.sg; michel.devoret@yale.edu

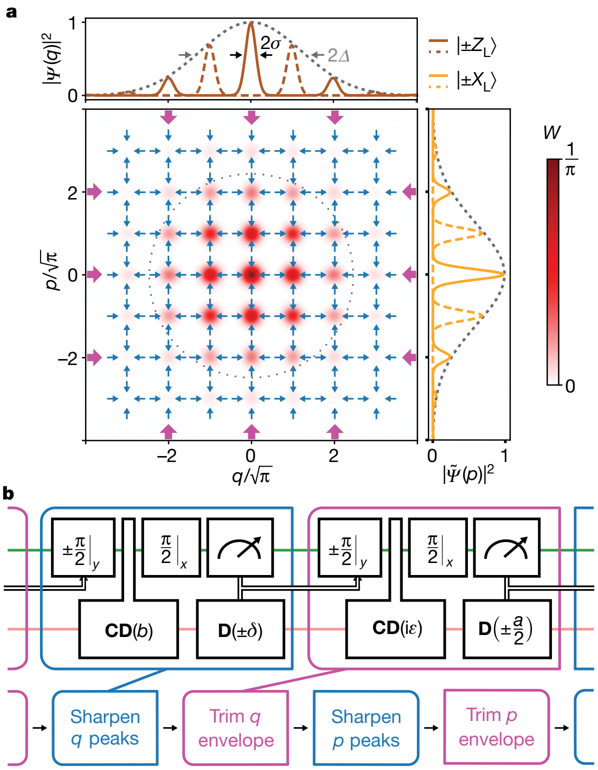


Fig. 1 | Quantum error correction protocol. **a**, Simulated Wigner function of the fully mixed logical state in a code defined by a width of $\sigma = 0.25$ for the peaks and of $\Delta = 1/(2\sigma) = 2$ for the normalizing envelope. Our QEC protocol prevents the squeezed peaks from spreading (blue arrows) and the overall envelope from extending (purple arrows). The side panels present the probability distributions of the $|\pm X_L\rangle$ and $|\pm Z_L\rangle$ states along each quadrature, which retain disjoint supports along q or p under stabilization. **b**, The full QEC protocol interleaves two peak-sharpening rounds and two envelope-trimming rounds to prevent spreading of the grid-state peaks and envelope in phase space (blue and purple arrows in **a**, respectively). In each round, a conditional displacement entangles the transmon (green line) and the storage oscillator (pink line). A subsequent measurement of the transmon controls the sign of a feedback shift of the oscillator and of a π/2 rotation resetting the transmon (bold black arrows). The peak-sharpening shift $\delta \approx 0.2$ maximizes the stabilizer value in the steady state, and the envelope-trimming conditional displacement of $\varepsilon \approx 0.2$ sets the width of the grid-state envelope (see Supplementary Figs. 10, 11), which is optimal given the experimental constraints.

Measurement of displacement operators

The expectation value of displacement operators $\mathbf{D}(\beta)$, such as the stabilizers and Pauli operators of the GKP code, are periodic functions of the generalized quadrature, $\mathbf{r} = -\mathrm{Re}(\beta)\mathbf{p} + i\mathrm{Im}(\beta)\mathbf{q}$. We measure these 'modular variables'[8,18,19] by effectively coupling the quadrature of an oscillator to the Pauli operator $\boldsymbol{\sigma}_z$ of an ancillary physical qubit. In our experiment, the oscillator is the fundamental mode of a reentrant coaxial microwave cavity made from bulk aluminium[20], which we call the storage mode, and the ancillary physical qubit is a transmon (see Supplementary Fig. 1). The storage mode has a single-photon lifetime of $T_s = 245$ μs (see Supplementary Fig. 5), and the transmon has energy and coherence lifetimes of $T_1 = 50$ μs and $T_{2E} = 60$ μs—measured with an echo sequence—and can be non-destructively measured in 700 ns via an ancillary low-quality-factor resonator (see Supplementary Table 1 and Supplementary Fig. 3). Interestingly, the desired coupling $\mathbf{r} \otimes \boldsymbol{\sigma}_z$ between the storage mode and the transmon can be effectively activated with microwave drives in the presence of the naturally present dispersive interaction[21], even with arbitrarily weak interaction strength.

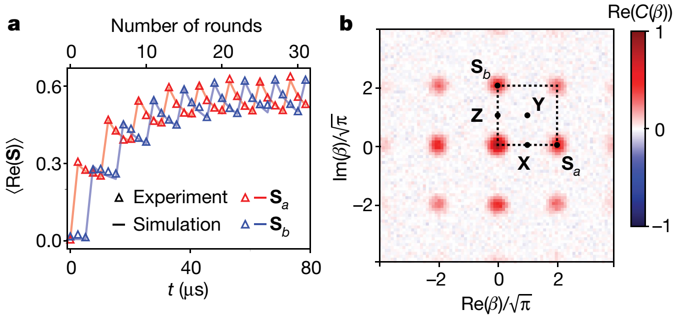


Fig. 2 | Square code in the steady state of the QEC protocol. **a**, Measured average value of the real part of the code stabilizers when turning on the QEC protocol from the vacuum state. Each stabilizer oscillates over a four-round period as a result of the periodicity of the QEC protocol, and the steady state is reached in about 20 rounds. **b**, Real part of the measured characteristic function of the storage mode in the steady state (after 200 rounds). Points corresponding to stabilizers and Pauli operators are indicated by black dots, and the dashed lines enclose an area of 4π.

Schematically, when the storage mode is displaced far from the origin of phase space, the dispersive interaction results in two quickly separating trajectories, each corresponding to a different transmon eigenstate. We employ this evolution within a sequence of fast storage displacements intertwined with transmon rotations to engineer an arbitrary conditional displacement in 1.1 μs, following the unitary evolution $\mathbf{CD}(\beta) = \exp[i(-\mathrm{Re}(\beta)\mathbf{p} + \mathrm{Im}(\beta)\mathbf{q})\frac{\boldsymbol{\sigma}_z}{2}]$ (see Supplementary Figs. 2, 4). This entangling gate can equivalently be viewed as a rotation of the transmon's Bloch vector around the $\boldsymbol{\sigma}_z$ axis by an angle that depends on the phase-space distribution of the storage mode. When applied to a transmon initialized on the equator of its Bloch sphere, it leads to $\langle \boldsymbol{\sigma}_x - i\boldsymbol{\sigma}_y \rangle = \langle \mathbf{D}(\beta) \rangle$ (ref. 8). Intuitively, given that the measurement of a displacement by β is a measurement of a quadrature modulo $2\pi/\beta$, the conditional displacement is such that two oscillator quadrature eigenstates separated by $2n\pi/\beta$ induce the same qubit rotation up to an integer number of turns n.

Conditional displacements embedded within a transmon Ramsey sequence enable the measurement of the code stabilizers and, therefore, lay at the heart of the QEC of GKP codes[22–24]. Conveniently, this sequence is also employed to obtain the expectation value of any displacement operator $\langle \mathbf{D}(\beta) \rangle$ for an arbitrary state of the storage oscillator. This leads to the state characteristic function $C(\beta)$, which is the two-dimensional Fourier transform of the Wigner function[25,26]. This complex-valued representation fully characterizes an arbitrary state. In our experiment we measure $\mathrm{Re}(C(\beta))$, which contains the information about the symmetric component of the Wigner function, to characterize the generated grid states presented in Figs. 2–4. The imaginary part, $\mathrm{Im}(C(\beta))$, contains information about the antisymmetric component of the Wigner function and is expected to take a uniform null value for the symmetric grid states that we consider. We verify this property at critical points.

Convergence to the GKP code manifold

We now derive a QEC protocol that employs the conditional displacement gate described earlier to protect finite-size grid states. Note that there exists an optimal width of the envelope Δ that results from a trade off: more extended grid states have better resolved peaks and are thus more robust against shifts, but are more sensitive to dissipation. Therefore, our protocol is designed: first, to keep the oscillator state probability distribution peaked in phase space at $q = 0 \bmod 2\pi/|a|$ and $p = 2\pi/|b|$; second, to prevent the overall envelope from drifting or expanding more than necessary. Given our experimental constraints, we work with a finite-size GKP code with envelope width $\Delta \approx 3.2$, chosen

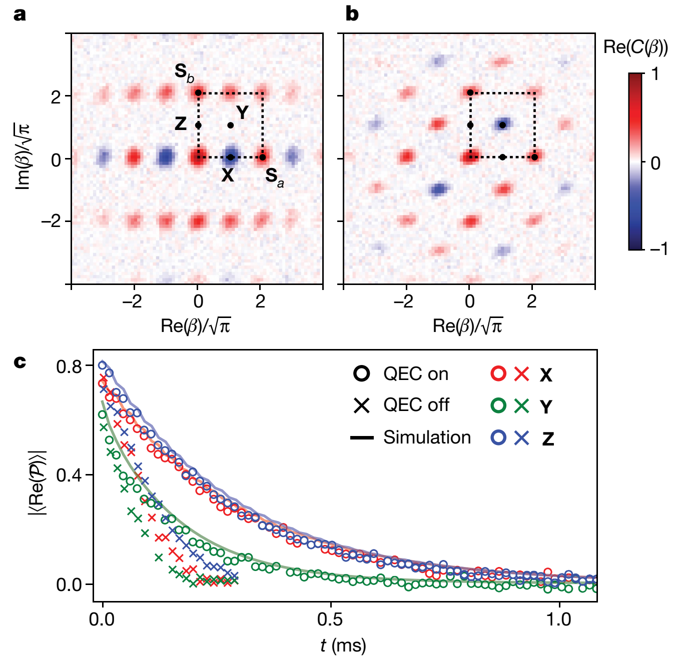


Fig. 3 | Initialization and coherence characterization of the logical qubit in the square encoding. **a**, Characteristic function of $|-X_L\rangle$, prepared in the steady state by applying a feedback **Z** gate conditioned on the outcome +1 of a first single-round Re(**X**) measurement, before heralding a higher-fidelity state on the outcome −1 of a second identical measurement. **b**, The same procedure, for $|-Y_L\rangle$. **c**, After preparing $|-X_L\rangle$, $|-Y_L\rangle$ or $|-Z_L\rangle$, the time decay of the real part of $\mathcal{P} = \mathbf{X}, \mathbf{Y}, \mathbf{Z}$, respectively, is measured while continuously applying the QEC protocol (on) or not (off). The QEC protocol extends the lifetime of the three Bloch vector components to $T_X = T_Z = 275$ μs and $T_Y = 160$ μs, and the results are quantitatively reproduced by master-equation simulations (lines).

to maximize the coherence time of the logical qubit (see Supplementary Fig. 9).

From the above discussion, maintaining the phase-space distribution peaked at the grid points involves mapping the stabilizers $\mathbf{S}_a$ or $\mathbf{S}_b$ onto the ancilla transmon with conditional displacements, and then performing actuating displacements based on transmon measurements. As the measurement of the transmon yields only a binary outcome, these steps are constructed to answer the simple questions of whether the grid has moved up or down (when measuring $\mathbf{S}_a$) and whether it has moved left or right (when measuring $\mathbf{S}_b$). After each measurement, we apply a fixed-length displacement in the direction opposite to that indicated by the answer (see Fig. 1). The combination of the back-action of the measurements and of our feedback sharpens the peaks of the grid states. Similar measurements of small displacement operators and feedback trim the envelope of the grid states to keep it from drifting and expanding (see Supplementary Fig. 10). The repeated action of this basic protocol forms a discrete-time Markovian sequence that leads to an effective dissipative force that pushes the state of the storage oscillator towards the code manifold, as depicted in Fig. 1a. This engineered dissipation counteracts the evolution due to noise, thereby inhibiting logical errors.

Starting from the ground state of the oscillator, we apply this protocol indefinitely, as summarized in Fig. 1b. In Fig. 2a we plot the measured average values of Re($\mathbf{S}_a$) and Re($\mathbf{S}_b$) after n correction rounds. The stabilizer values increase rapidly to converge to a steady state in about 20 rounds. In addition to this trend, the mean value of each stabilizer oscillates over a period of four rounds by increasing to 0.62 when the peaks are sharpened in the corresponding phase-space quadrature, and then decays to 0.5 over the next three rounds. Beyond this periodic oscillation, the stabilizers do not evolve over hundreds of rounds (not shown), which indicates that our protocol has entered a steady state. The characterization of this steady state can now reveal whether it corresponds to the desired GKP manifold.

We plot the real part of the characteristic function of the steady state after 200 rounds in Fig. 2b. This state is a maximally mixed state of the logical qubit, as can be seen from the null value of the points corresponding to the three logical Pauli operators. Note that this characteristic function representation is the Fourier conjugate of the theoretical Wigner representation given in Fig. 1a. However, the two are similar for grid states because the Fourier transform of a grid is itself a grid. Our results are quantitatively reproduced by master-equation simulations (lines in Fig. 2a), the parameters of which are all calibrated independently. From these simulations, we estimate that the squeezing of the peaks of the generated grid states oscillates between 7.4 dB and 9.5 dB in the steady state—close to the level required for fault-tolerant quantum computation[27–29]—and the average photon number oscillates between 8.6 and 10.2.

Logical qubit initialization

Once the oscillator has reached its steady state, it is in the code manifold, and we initialize the logical qubit by replacing one of the QEC rounds with a measurement of **X**, **Y** or **Z**. To measure the logical Pauli operators, we first prepare the transmon in $|+x\rangle$ and then apply the conditional displacement **CD**(β) with $\beta = a/2$, $(a+b)/2$ or $b/2$, respectively. After the sequence, $\langle \boldsymbol{\sigma}_x - \mathrm{i}\boldsymbol{\sigma}_y\rangle = \langle \mathbf{X}\rangle$, $\langle \mathbf{Y}\rangle$ or $\langle \mathbf{Z}\rangle$, and a subsequent $\boldsymbol{\sigma}_x$ readout of the transmon with outcome ±1 heralds the preparation of the approximately orthogonal states $|\pm X_L\rangle$, $|\pm Y_L\rangle$ or $|\pm Z_L\rangle$ up to a re-centring displacement (see Supplementary Fig. 9).

However, because **X**, **Y** or **Z** differ from the Pauli operators of the finitely squeezed code that we consider, the sequence described above results in a readout of the logical qubit with non-unit fidelity and in an imperfect initialization. Fortunately, when this sequence is followed by a few QEC rounds projecting the generated state back onto the code manifold, this readout is non-demolition for the target logical state and can be repeated to increase its fidelity (see Supplementary Information). In Fig. 3a (Fig. 3b) we show the characteristic function of the storage state obtained when two **X** (**Y**) measurements, separated by four QEC rounds, yield the same outcome. The expectation values of the Pauli operators in these two cases are $\langle \mathrm{Re}(\mathbf{X})\rangle = -0.8$ and $\langle \mathrm{Re}(\mathbf{Y})\rangle = -0.63$, respectively. We emphasize here that these values do not reflect the preparation fidelity to the finitely squeezed logical states $|-X_L\rangle$ and $|-Y_L\rangle$, and the prepared state is as close (within experimental uncertainties) to the target state as allowed by the imperfect code correction (see Supplementary Information). The same methods are applied to prepare eigenstates of other Pauli operators (data not shown) and can be modified to prepare non-Clifford states (see Supplementary Fig. 13). In particular, the characteristic function of the $|-Z_L\rangle$ state is the same as that of $|-X_L\rangle$ rotated by 90° (see Supplementary Fig. 7).

Coherence of the error-corrected logical qubit

To test the error-correction performance of our protocol, we prepare one of the logical states $|-X_L\rangle$, $|-Y_L\rangle$ or $|-Z_L\rangle$, and compare the decay of the mean value of the real part of the corresponding operator $\mathcal{P} = \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ in time when performing QEC (open circles in Fig. 3b) and when not (crosses in Fig. 3b). In all three cases, our protocol extends the coherence of the logical qubit. We extract the coherence times of the error-corrected qubit $T_X = T_Z = 275$ μs and $T_Y = 160$ μs. The shorter coherence time of the **Y** Pauli operator, also visible in the uncorrected case, is expected, because the distance in phase space from the probability peaks of the $|+Y_L\rangle$ state to those of the $|-Y_L\rangle$ state is shorter by $\sqrt{2}$ than in the case of $|\pm X_L\rangle$ and $|\pm Z_L\rangle$. Therefore, diffusive shifts in phase space induced by photon dissipation cause more flips of the **Y** component

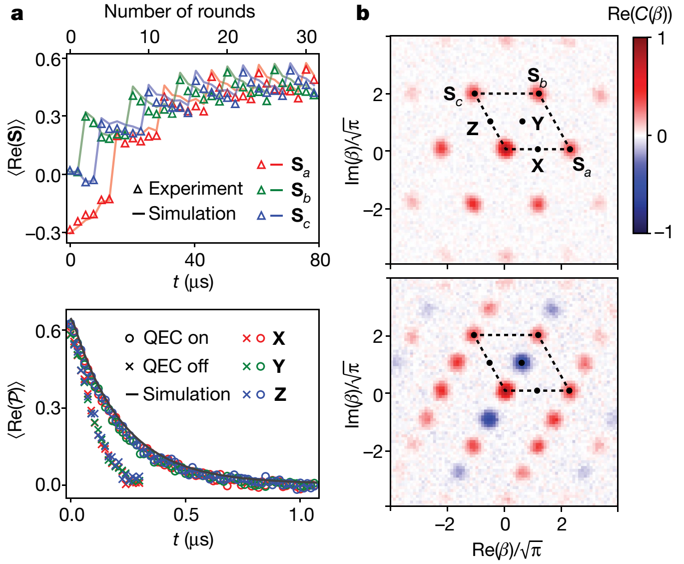


Fig. 4 | Convergence to the code manifold, state preparation and coherence in the hexagonal code. a, The grid-state peaks and envelope are sequentially sharpened and trimmed along three directions. When turning on our protocol from the ground state of the oscillator, the real part of the expectation values of the stabilizers oscillates every six rounds and increases to rapidly reach a steady state. **b**, After 200 rounds, the oscillator state is a fully mixed logical state that reveals the code structure (top). An eigenstate of a Pauli operator, such as $|-Y_\mathrm{L}\rangle$ (bottom), can be prepared by a single-round measurement of Re(**Y**), followed by a feedback displacement. **c**, Owing to the code symmetry, the decay of the logical Bloch vector is isotropic. An exponential fit (black line) indicates a lifetime of 205 μs, enhanced by QEC.

of the logical qubit Bloch vector. Master-equation simulations reproduce these results quantitatively.

Hexagonal code

We executed a variant of the square code of Fig. 1 known as the hexagonal code, in which the decay times of all three Pauli operators are equal by symmetry. In general, a two-dimensional grid-state code is defined as the common eigenspace of any two commuting stabilizers $\mathbf{S}_a = \mathbf{D}(a)$ and $\mathbf{S}_b = \mathbf{D}(b)$, as long as $\mathrm{Im}(a^*b) = 4\pi$. Geometrically, this condition implies that the magnitude of the cross-product of the two vectors representing these stabilizers corresponds to an area of 4π (see Figs. 2b, 4b, Supplementary Fig. 12). In the hexagonal GKP code[2], we have $b = a\exp(\mathrm{i}\frac{\pi}{3})$, which respects the above area condition for $a = \sqrt{(8\pi/\sqrt{3})}$. The Pauli operators correspond to displacements of equal length, $\mathbf{X} = \mathbf{D}(a/2)$, $\mathbf{Y} = \mathbf{D}(b/2)$ and $\mathbf{Z} = \mathbf{D}(c/2)$ with $c = a\exp(\mathrm{i}\frac{2\pi}{3})$. For symmetry reasons, we also define a third stabilizer, $\mathbf{S}_c = \mathbf{Z}^2 = \mathbf{D}(c)$, that commutes with the two others.

We perform QEC on this code by adapting the protocol described in section 'Convergence to the GKP code manifold'. Here, measurement of the three hexagonal stabilizers, followed by small corrective feedback displacements, sharpens the peaks along three different directions. These steps are interleaved with the measurement of three short displacement operators, which trim the envelope. When applying this protocol on the storage mode initialized in the ground state, the mean values of the stabilizers oscillate every six rounds as each of these displacement operators is measured in turn, and rapidly converge to a stationary regime in which their values oscillate between 0.4 and 0.55 (see Fig. 4a). We measure the real part of the characteristic function of the fully mixed logical state reached after 200 rounds, which reveals the hexagonal structure of the code (Fig. 4b). Again, master-equation simulations reproduce these results quantitatively and indicate that the generated grid states are characterized by the same squeezing for the peaks as in the square encoding (between 7.5 dB and 9.5 dB in the steady state). Note that the temporary negative value of Re($\mathbf{S}_a$) registered at short times originates from the programming of the feedback algorithm on the fast FPGA (field-programmable gate array) board: the oscillator state gets shifted at the beginning of the sequence, which is included in the simulations.

We prepare the logical qubit in an eigenstate of each Pauli operator with a single-round measurement of Re(**X**), Re(**Y**) or Re(**Z**). In Fig. 4b we show the measured characteristic function of the $|-Y_\mathrm{L}\rangle$ state. We note that the characteristic functions of $|-X_\mathrm{L}\rangle$ and $|-Z_\mathrm{L}\rangle$ are equal to that of $|-Y_\mathrm{L}\rangle$ but rotated by ±60° (see Supplementary Fig. 8). Finally, we characterize the coherence of the error-corrected logical qubit by measuring the decay of the Pauli operator mean values in time. As expected, the decoherence of the logical qubit is now isotropic and considerably extended compared to the uncorrected case, with coherence times of $T_X = T_Y = T_Z = 205$ μs.

Logical errors and outlook

The coherence of the logical qubit is limited by two factors. First, the duration of the error-correction rounds, despite being a factor of 100 shorter than the storage-mode single-photon lifetime, is not negligible. The transmon readout and its processing using the FPGA accounts for about half of this duration, and the conditional displacement gate accounts for the other half. Although the gate speed is limited in this implementation, alternative implementations could result in faster gates[30]. The second factor limiting the coherence of the logical qubit is transmon errors. Among these, $\boldsymbol{\sigma}_\mathbf{z}$ errors (phase-flips) commute with the storage–transmon interaction Hamiltonian and thus do not propagate to the logical qubit (see Supplementary Information). On the other hand, the $\boldsymbol{\sigma}_\mathbf{x}$ and $\boldsymbol{\sigma}_\mathbf{y}$ transmon errors (bit-flips), as well as excitations to the higher excited states of the transmon (see Supplementary Fig. 6), propagate to the logical qubit as they lead to random displacements of the storage mode. Simulations indicate that bit-flips of the transmon and the finite correction rate each account for about half of the error rate of the logical qubit (see Supplementary Table 2).

The coherence of the logical qubit could be further improved by replacing the transmon with a noise-biased ancillary qubit[31–33] and by using a superconducting cavity with a larger quality factor[20]. This multipronged effort at improving the GKP code using superconducting circuits will be particularly rewarding because fault-tolerant single- and multi-qubit Clifford gates can be implemented in a straightforward way[2,34], and such logical qubits can be embedded in further layers of protection[27–29,35].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2603-3.

1. Shor, P. Fault-tolerant quantum computation. In *Proc. 37th Conf. Foundations of Computer Science* 56–65 (IEEE, 1996).
2. Gottesman, D., Kitaev, A. & Preskill, J. Encoding a qubit in an oscillator. *Phys. Rev. A* **64**, 012310 (2001).
3. Travaglione, B. & Milburn, G. J. Preparing encoded states in an oscillator. *Phys. Rev. A* **66**, 052322 (2002).
4. Pirandola, S., Mancini, S., Vitali, D. & Tombesi, P. Continuous variable encoding by ponderomotive interaction. *Eur. Phys. J. D* **37**, 283–290 (2006).
5. Vasconcelos, H. M., Sanz, L. & Glancy, S. All-optical generation of states for "encoding a qubit in an oscillator". *Opt. Lett.* **35**, 3261–3263 (2010).
6. Terhal, B. & Weigand, D. Encoding a qubit into a cavity mode in circuit QED using phase estimation. *Phys. Rev. A* **93**, 012315 (2016).
7. Motes, K. R., Baragiola, B. Q., Gilchrist, A. & Menicucci, N. C. Encoding qubits into oscillators with atomic ensembles and squeezed light. *Phys. Rev. A* **95**, 053819 (2017).

8. Flühmann, C., Negnevitsky, V., Marinelli, M. & Home, J. P. Sequential modular position and momentum measurements of a trapped ion mechanical oscillator. *Phys. Rev. X* **8**, 021001 (2018).
9. Flühmann, C. et al. Encoding a qubit in a trapped-ion mechanical oscillator. *Nature* **566**, 513–517 (2019).
10. Waldherr, G. et al. Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).
11. Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).
12. Cramer, J. et al. Repeated quantum error correction on a continuously encoded qubit by real-time feedback. *Nat. Commun.* **7**, 11526 (2016).
13. Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441–445 (2016).
14. Hu, L. et al. Quantum error correction and universal gate set operation on a binomial bosonic logical qubit. *Nat. Phys.* **15**, 503–508 (2019).
15. Albert, V. V. et al. Performance and structure of single-mode bosonic codes. *Phys. Rev. A* **97**, 032346 (2018).
16. Noh, K., Albert, V. V. & Jiang, L. Quantum capacity bounds of gaussian thermal loss channels and achievable rates with Gottesman–Kitaev–Preskill codes. *IEEE Trans. Inf. Theory* **65**, 2563–2582 (2019).
17. Cahill, K. E. & Glauber, R. J. Ordered expansions in boson amplitude operators. *Phys. Rev.* **177**, 1857–1881 (1969).
18. Aharonov, Y., Pendleton, H. & Petersen, A. Modular variables in quantum theory. *Int. J. Theor. Phys.* **2**, 213–230 (1969).
19. Popescu, S. Dynamical quantum non-locality. *Nat. Phys.* **6**, 151–153 (2010).
20. Reagor, M. et al. Quantum memory with millisecond coherence in circuit QED. *Phys. Rev. B* **94**, 014506 (2016).
21. Wallraff, A. et al. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
22. Kitaev, A. Y. Quantum measurements and the abelian stabilizer problem. Preprint at https://arxiv.org/abs/quant-ph/9511026 (1995).
23. Svore, K. M., Hastings, M. B. & Freedman, M. Faster phase estimation. *Quant. Inf. Comp.* **14**, 306–328 (2013).
24. Weigand, D. J. & Terhal, B. M. Generating grid states from Schrödinger-cat states without postselection. *Phys. Rev. A* **97**, 022341 (2018).
25. Haroche, S. & Raimond, J.-M. *Exploring the Quantum* (Oxford Univ. Press, 2006).
26. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer Science & Business Media, 2007).
27. Glancy, S. & Knill, E. Error analysis for encoding a qubit in an oscillator. *Phys. Rev. A* **73**, 012325 (2006).
28. Fukui, K., Tomita, A., Okamoto, A. & Fujii, K. High-threshold fault-tolerant quantum computation with analog quantum error correction. *Phys. Rev. X* **8**, 021054 (2018).
29. Vuillot, C., Asasi, H., Wang, Y., Pryadko, L. P. & Terhal, B. M. Quantum error correction with the toric Gottesman–Kitaev–Preskill code. *Phys. Rev. A* **99**, 032344 (2019).
30. Touzard, S. et al. Gated conditional displacement readout of superconducting qubits. *Phys. Rev. Lett.* **122**, 080502 (2019).
31. Puri, S. et al. Stabilized cat in driven nonlinear cavity: a fault-tolerant error syndrome detector. *Phys. Rev. X* **9**, 041009 (2019).
32. Grimm, A. et al. The Kerr-cat qubit: stabilization, readout, and gates. *Nature* **584**, 205–209 (2020).
33. Shi, Y., Chamberland, C. & Cross, A. W. Fault-tolerant preparation of approximate GKP states. Preprint at https://arxiv.org/abs/1905.00903 (2019).
34. Gao, Y. Y. et al. Programmable interference between two microwave quantum memories. *Phys. Rev. X* **8**, 021073 (2018).
35. Baragiola, B. Q., Pantaleoni, G., Alexander, R. N., Karanjai, A. & Menicucci, N. C. All-Gaussian universality and fault tolerance with the Gottesman–Kitaev–Preskill code. Preprint at https://arxiv.org/abs/1903.00012 (2019).

Data availability

The experimental data and numerical simulations presented here are available from the corresponding authors upon request.

Acknowledgements We thank C. Flühmann, J. Home, S. Girvin, L. Jiang and K. Noh for discussions and M. Rooks for fabrication assistance. M.M. thanks the Yale Quantum Institute for hosting him while he was collaborating on this project. The use of facilities was supported by YINQE and the Yale SEAS cleanroom. This research was supported by ARO under grant number W911NF-18-1-0212 and ARO grant number W911NF-16-1-0349.

Author contributions P.C.-I., A.E. and S.T. designed and performed the experiment and analysed the data. E.Z.-G. N.E.F., V.V.S., P.R., S.S., R.J.S. and L.F. contributed to the experimental apparatus, and S.P. and M.M. provided theoretical support. M.H.D. supervised the project. P.C.-I., A.E., S.T. and M.H.D. wrote the manuscript. All authors provided suggestions for the experiment, discussed the results and contributed to the manuscript.

Competing interests L.F., R.J.S. and M.H.D. are founders of QCI. L.F. and R.J.S. are shareholders of QCI. All authors, except A.E. and E.Z.G., are inventors of patents (USA, Japan and Singapore) related to the subject.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2603-3.

Correspondence and requests for materials should be addressed to P.C.-I., S.T. or M.H.D.

Peer review information *Nature* thanks Barbara Terhal and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Observation of superconducting diode effect

https://doi.org/10.1038/s41586-020-2590-4

Received: 14 March 2020

Accepted: 23 June 2020

Fuyuki Ando[1], Yuta Miyasaka[1], Tian Li[1], Jun Ishizuka[2], Tomonori Arakawa[3,4], Yoichi Shiota[1], Takahiro Moriyama[1], Youichi Yanase[2] & Teruo Ono[1,4] ✉

Nonlinear optical and electrical effects associated with a lack of spatial inversion symmetry allow direction-selective propagation and transport of quantum particles, such as photons[1] and electrons[2–9]. The most common example of such nonreciprocal phenomena is a semiconductor diode with a p–n junction, with a low resistance in one direction and a high resistance in the other. Although the diode effect forms the basis of numerous electronic components, such as rectifiers, alternating–direct-current converters and photodetectors, it introduces an inevitable energy loss due to the finite resistance. Therefore, a worthwhile goal is to realize a superconducting diode that has zero resistance in only one direction. Here we demonstrate a magnetically controllable superconducting diode in an artificial superlattice $[\mathrm{Nb/V/Ta}]_n$ without a centre of inversion. The nonreciprocal resistance versus current curve at the superconducting-to-normal transition was clearly observed by a direct-current measurement, and the difference of the critical current is considered to be related to the magnetochiral anisotropy caused by breaking of the spatial-inversion and time-reversal symmetries[10–13]. Owing to the nonreciprocal critical current, the $[\mathrm{Nb/V/Ta}]_n$ superlattice exhibits zero resistance in only one direction. This superconducting diode effect enables phase-coherent and direction-selective charge transport, paving the way for the construction of non-dissipative electronic circuits.

Nonreciprocal charge transport is important for the wide use of electronic components, such as rectifiers, alternating–direct-current converters and photodetectors. In 1874, Braun discovered rectification in a metal–semiconductor contact[2], heavily influencing the development of semiconductor devices. In modern condensed matter physics, Rikken et al.[1,3–5] demonstrated that nonlinear optical and electrical responses can generally be achieved when both spatial inversion and time-reversal symmetry are broken in a system, a phenomenon described as magnetochiral anisotropy[1–9]. In a Rashba system, where the spatial inversion is uniaxially broken along the z axis, the spin–orbit interaction causes spin-dependent band splitting[14–16]. The spin σ and the wavevector k are required to be orthogonal in the x–y plane and the electrons with $+k$ and $-k$ have opposite spin directions ($\sigma(+k) = -\sigma(-k)$). For example, if we apply a magnetic field $\mathbf{B}_y$ to break time-reversal symmetry along the y axis, the electrons with wavevectors $+k_x$ and $-k_x$ come to have non-equivalent energy depending on whether the spin $\sigma(\pm k_x)$ is parallel or antiparallel to the magnetic field $\mathbf{B}_y$. This results in a nonlinear response of current $\mathbf{I}_x$ proportional to the square of the electric field along the x axis under magnetic field $\mathbf{B}_y$. This nonreciprocal effect, induced by the Rashba spin–orbit interaction and magnetic field, is expressed in the form of current-dependent resistance as shown in equation (1).

$$R = R_0(1 + \gamma(\mathbf{B} \times \mathbf{z}) \cdot \mathbf{I}) \qquad (1)$$

where γ is the coefficient of magnetochiral anisotropy depending on the Rashba spin–orbit interaction. The nonlinear resistance is considered to be the perturbation to the linear resistance R_0 that is generally scaled by the kinetic energy of the electrons. Therefore, the magnitude of γ is typically very tiny in normal conductors because the Rashba spin–orbit interaction and magnetic energy are much smaller than the kinetic energy of the electrons, that is, the Fermi energy E_F.

Recently, nonreciprocal charge transport in superconductors has attracted considerable interest because the nonlinear resistance was found to be remarkably enhanced in the superconducting fluctuation region compared to the normal conducting state[10–13,17]. This trend can be explained by the replacement of the energy denominator in the second term of equation (1) from E_F (in electronvolts) to the energy gap (in millielectronvolts) in the superconductors[10,13]. The discovery strongly suggests the potential for directional transport of superconducting current. However, for low-dimensional superconductors such as $\mathrm{MoS_2}$ (ref. [10]), $\mathrm{WS_2}$ (ref. [11]) and $\mathrm{Bi_2Te_3}$/FeTe (ref. [12]), the rectification ratio is not sufficient for implementation in devices because the linear resistance R_0 gradually decreases during the superconducting transition and is orders of magnitude larger than the nonlinear one. Consequently, there still remains the need to realize a superconducting diode that has zero resistance in only one direction.

Here we fabricate a noncentrosymmetric superlattice[18,19] by stacking three kinds of superconducting elements—niobium, vanadium and tantalum—repeatedly, and observe a superconducting diode effect controlled by a magnetic field (Fig. 1). We clearly observed nonreciprocity in the resistance–current (R–I) curve, specifically the critical current, by direct-current (d.c.) measurement (Fig. 2). Furthermore, the nonlinear resistance unique to the Rashba superconductor was also observed in the resistive superconducting fluctuation region. This means that the Rashba superconductivity can be accessed using an artificially

---

[1]Institute for Chemical Research, Kyoto University, Kyoto, Japan. [2]Department of Physics, Graduate School of Science, Kyoto University, Kyoto, Japan. [3]Graduate School of Science, Osaka University, Osaka, Japan. [4]Center for Spintronics Research Network, Graduate School of Engineering Science, Osaka University, Osaka, Japan. ✉e-mail: ono@scl.kyoto-u.ac.jp

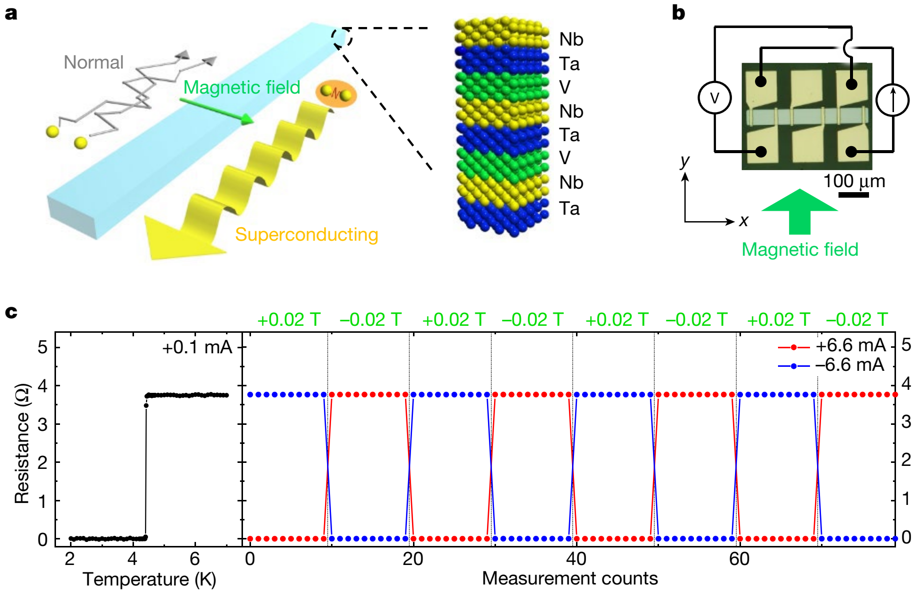


Fig. 1 | Demonstration of the magnetically controllable superconducting diode. a, Schematic images of the superconducting diode controlled by an external magnetic field and the artificial [Nb/V/Ta]$_n$ superlattice, in which the global inversion symmetry is broken along the direction of stacking. When the directions of the current, the magnetic field and inversion symmetry breaking are orthogonal to one other, the Cooper pairs can flow in only one direction. **b**, Photomicrograph of the processed device and the measurement setup with the definitions of electric current and magnetic field. **c**, Temperature dependence of the sheet resistance of the [Nb(1.0 nm)/V(1.0 nm)/Ta(1.0 nm)]$_{40}$ film and alternating switching between the superconducting and normal conducting states by changing the sign of the applied current or magnetic field at 4.2 K.

engineered three-dimensional superlattice. The nonreciprocal critical current presented here can be considered to be a consequence of the magnetochiral anisotropy in the Rashba superconductor.

The superlattice [Nb (1.0 nm)/V (1.0 nm)/Ta (1.0 nm)]$_{40}$, which was epitaxially grown on the MgO (100) substrate with well defined periodic interfaces[20], was fabricated into a wire structure for four-terminal measurements (see Methods and Fig. 1b). An external magnetic field was applied by a superconducting magnet in a Physical Property Measurement Systems (PPMS-3) chamber, whose direction was in the plane of the film and perpendicular to the flowing current **I**. Here, the current and magnetic field directions are defined as the x and y axes, respectively.

First, we measured the temperature dependence of the sheet resistance with a small d.c. current $I = +0.1$ mA to determine the critical temperature T_c. As shown in Fig. 1c, we found that the T_c is 4.41 K and the normal resistance is around 3.75 Ω at low temperature. The alternating switching between the superconducting and normal conducting states (Fig. 1c) was demonstrated as follows: after setting the magnetic field at +0.02 T (or −0.02 T) in the $+y$ direction, the sheet resistance was continuously measured with the positive d.c. current $I = +6.6$ mA at 4.2 K, slightly below T_c. Then, whereas zero resistance (about 0.0017 Ω) was obtained with the positive magnetic field, normal resistance (3.76 Ω) was obtained when the magnetic field reversed. When we applied a negative d.c. current $I = -6.6$ mA, the switching behaviour reversed. This result strongly indicates that the superconducting and normal conducting states are fully switched depending on the sign of the magnetic field and the current. Two important features are: a rectification ratio over 2,000, comparable to those of typical semiconductor diodes, and nonreciprocity that can be easily controlled by a small magnetic field.

To investigate the superconducting diode effect in detail, we measured the d.c. current dependence of the sheet resistance under various magnetic fields and temperatures (Fig. 2). In the typical measurement results at 4.2 K (Fig. 2a), we notice that the R–I curves show a jump at different currents depending on whether the applied d.c. currents are positive ($+I$) or negative ($-I$). The sharpness of the phase transition, perhaps owing to the three-dimensionality of the [Nb/V/Ta]$_n$ superlattice compared to other low-dimensional superconductors, allows switching between the superconducting and normal conducting states (Fig. 1c). Here, the midpoint of the R–I curve is defined as the critical current I_c, and the I_c values under the various magnetic fields are plotted for both positive ($+I$) and negative ($-I$) currents (Fig. 2b). The upper critical field B_{c2} is estimated to be 0.2 T from the I_c curves, which means the diode effect demonstrated in Fig. 1c can be controlled by a tenth or less of the B_{c2}. These two curves clearly suggest that the sign of the nonreciprocal components in I_c is uniquely determined by the relative angle between the current and magnetic field directions, where I_c increases when the magnetic field is directed left of the current and decreases when directed right of the current. Next, we investigated the temperature dependence of the nonreciprocal critical current to characterize its behaviour. Here, the nonreciprocal component ΔI_c is defined as in equation (2).

$$\Delta I_c = I_c(+I) - I_c(-I) \qquad (2)$$

The magnetic field dependence of the ΔI_c was investigated in the range 2.0–4.35 K (Fig. 2c). For each temperature, the results where the magnetic field was swept forward ($+y$) and backward ($-y$) along the y axis exhibit an antisymmetric behaviour with regard to the magnetic field; this confirms that the ΔI_c is intrinsically determined by the magnetic field. We find that as the temperature increases towards the T_c, the ΔI_c clearly appears and subsequently shrinks, which resembles the behaviour of the nonreciprocal charge transport in MoS_2 (ref. [10]), WS_2 (ref. [11]) and Bi_2Te_3/FeTe (ref. [12]). To understand the temperature dependence of the diode effect, it will be desirable to develop a microscopic theory of the critical current.

We performed an alternating current (a.c.) harmonic measurement to discover the mechanism of the nonreciprocal critical current by comparing with the nonlinear resistance in equation (1). Once again, the nonreciprocal nature induced by Rashba spin–orbit interaction appears in the form of current-dependent resistance under the magnetic field B. To distinguish the linear and nonlinear resistances, the first- and second-harmonic sheet resistances (R_ω and $R_{2\omega}$) were measured using a lock-in amplifier under the application of an a.c. current

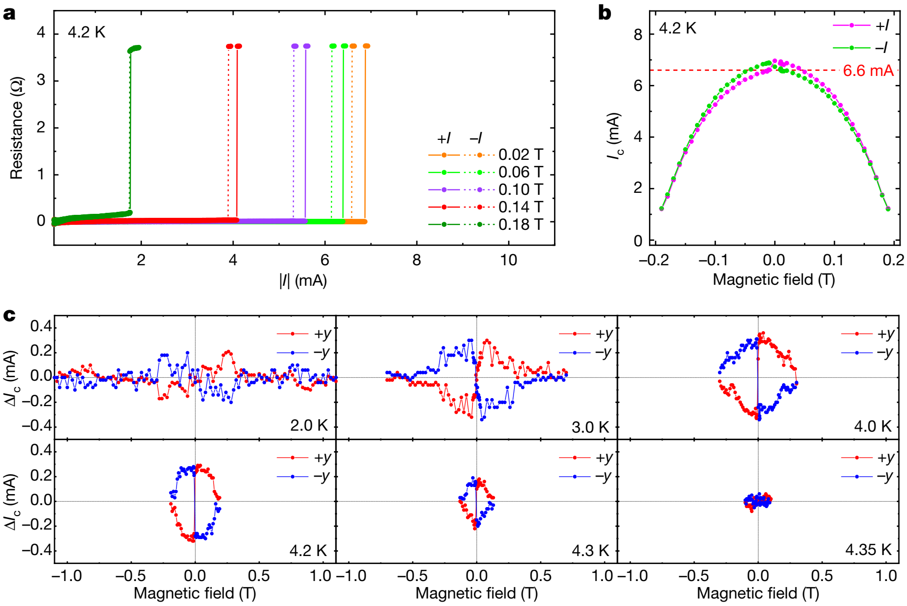


Fig. 2 | Asymmetric *R*–*I* curves and the nonreciprocal critical currents in the [Nb/V/Ta]$_n$ superlattice. a, Current dependences of the sheet resistance under various magnetic fields for both positive and negative currents at 4.2 K. **b**, Nonreciprocal critical current I_c as a function of the magnetic field. The red dashed line indicates a current of 6.6 mA, corresponding to the current amplitude, demonstrating the superconducting diode effect shown in Fig. 1. **c**, The nonreciprocal component of the critical current ΔI_c plotted as a function of the magnetic field at various temperatures. As the temperature increases towards the T_c, the ΔI_c clearly appears and subsequently shrinks.

with an amplitude of 2.0 mA and a frequency of 503 Hz. The other measurement configurations, such as the wire structure and the magnetic field, are the same as those of the d.c. measurement (Fig. 2b). Here, R_ω corresponds to the linear resistance R_0, which is independent of the current magnitude. In contrast, $R_{2\omega}$ represents the second-order resistance, which is proportional to the current and the magnetic field ($R_{2\omega} = \frac{R_0}{2}\gamma BI$). Both R_ω and $R_{2\omega}$ were measured while sweeping the magnetic field at 4.2 K (Fig. 3a). $R_{2\omega}$ is greatly enhanced during the transitions and antisymmetric with respect to the magnetic field when the magnetic field is orthogonal to the current (±*y* direction), whereas it is negligibly small when the magnetic field is set parallel to the current (+*x* direction). This direction-dependent $R_{2\omega}$ is characteristic of the Rashba superconductors. To our knowledge, this is the first direct evidence of the Rashba superconductivity in a three-dimensional superlattice. We next measured the temperature dependence of the $R_{2\omega}$ signals, to compare with that of the ΔI_c (Fig. 3b). The $R_{2\omega}$ signals observed at each temperature are enhanced around the critical field, perhaps because the nonreciprocal charge transport may be caused by superconducting fluctuations[10,13,21].

The temperature dependence of the γ values, which are calculated as $\gamma = \frac{2R_{2\omega}}{R_\omega BI}$ (see equation (1)), is plotted in Fig. 4. Here, we adopted the maximum $R_{2\omega}$ for each temperature and the corresponding R_ω and magnetic field. The notable observations are that the γ value increases in the vicinity of T_c and reaches $\gamma \approx 550\ \mathrm{T^{-1}\,A^{-1}}$ at its maximum. These trends are entirely consistent with those in other low-dimensional systems. The magnetochiral anisotropy of the [Nb/V/Ta]$_n$ superlattice naturally leads to nonreciprocal critical current; this was also clearly observed in the vicinity of T_c.

Finally, we discuss the possible origin of the magnetochiral anisotropy in the [Nb/V/Ta]$_n$ superlattice from a theoretical point of view. A first-principles calculation was performed to identify the band structure of a [Nb/V/Ta]$_n$ superlattice, where two-atomic-layer slabs of body-centred cubic Nb, V, and Ta were repeatedly stacked five times (see Methods). Indeed, we noticed the Rashba spin–orbit interaction-induced energy splitting near E_F (Extended Data Fig. 1), which was estimated to be $E_R = 10$ meV at maximum (about 1% of E_F). The Rashba splitting is induced by the artificially engineered superlattice structure of intrinsically centrosymmetric metals. Thus, the Rashba splitting, which causes nonreciprocal charge transport in the [Nb/V/Ta]$_n$ superlattice (Fig. 3), is also verified theoretically, and can be controlled by the superlattice structures. Our first-principles calculation reveals that electronic states near E_F are composed of strongly hybridized orbitals of Ta, Nb and V atoms. Such orbitals are affected by asymmetric heterostructures, and a combination with a large atomic spin–orbit interaction of Nb and Ta atoms generates a sizeable Rashba spin–orbit interaction. Here, an essential feature of the Rashba superconductor is the mixing state of the spin singlet and triplet pairings[22–25]. According to the predictions from Wakatsuki and Nagaosa[21], one of the mechanisms of magnetochiral anisotropy in Rashba superconductors could be the superconducting fluctuation with the mixed spin-singlet and spin-triplet pairings. If we adopt this

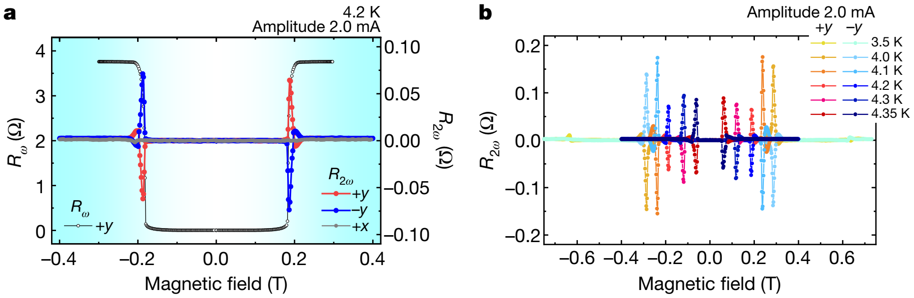


Fig. 3 | Nonreciprocal charge transport during the superconducting transition in the [Nb/V/Ta]$_n$ superlattice. a, Magnetic field dependence of first- (R_ω) and second-harmonic ($R_{2\omega}$) sheet resistances at 4.2 K. The white and blue shadings indicate the superconducting and normal conducting regions, respectively. $R_{2\omega}$ values are clearly enhanced when the directions of the current and magnetic field are orthogonal. **b**, Temperature dependence of the antisymmetric second-harmonic sheet resistances.

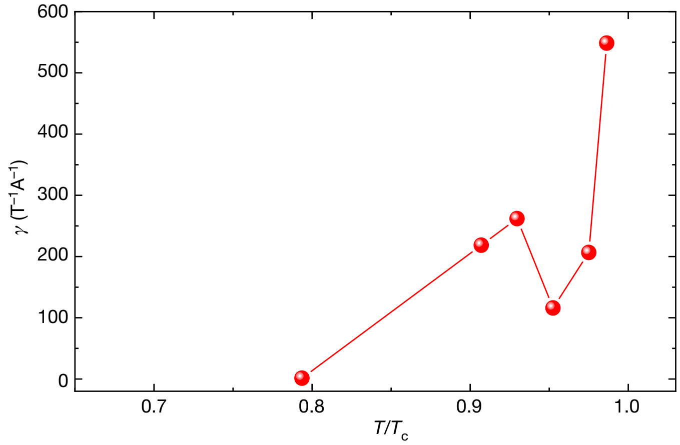


Fig. 4 | Magnetochiral anisotropy of the [Nb/V/Ta]$_n$ superlattice. The coefficient of magnetochiral anisotropy γ determined from $R_{2\omega}/R_\omega$ as a function of temperature. Although a dip appears reflecting the small $R_{2\omega}$ at 4.2 K and 4.3 K (Fig. 3b), the plot roughly shows the trend of γ increasing in the vicinity of T_c.

mechanism, the spin-triplet Cooper pairs induced by Rashba splitting have essential roles in the nonreciprocal charge transport and critical current in the [Nb/V/Ta]$_n$ superlattice. Since the ratio of spin-triplet to spin-singlet pair amplitudes is of the same order of magnitude as E_R/E_F (refs. [23,26,27]), we expect parity-mixing with a spin-triplet component of approximately 1%. However, not only the spin-triplet pair amplitude but also the spin-triplet pairing interaction is essential for the nonreciprocal transport[21]. The pairing interaction depends on the details of the system[23,26–29]. Moreover, this model neglects the interband scatterings due to multiband structure and magnetochiral anisotropy originating from other mechanisms, such as vortex dynamics. Actually, vortices are likely to enter along the field direction because the coherence length of 13 nm (see Methods and Extended Data Fig. 3) is much smaller than the thickness of 120 nm. These problems remain to be solved before we can obtain a more accurate estimate of the pairing compositions in the [Nb/V/Ta]$_n$ superlattice. We note that the vortex dynamics associated with the Berezinskii–Kosterlitz–Thouless transition[13,30] characteristic of two-dimensional superconductors is unlikely to be the origin of the nonreciprocal transport in the [Nb/V/Ta]$_n$ superlattice, because the film is much thicker than the coherence length. Our theoretical understanding of the nonreciprocal charge transport and critical current is as yet far from complete.

In conclusion, we have demonstrated a magnetically controllable superconducting diode in an artificial [Nb/V/Ta]$_n$ superlattice. The nonreciprocal R–I curves, revealing the nonreciprocal critical current, were observed using d.c. measurements, and should be related to the magnetochiral anisotropy effect induced by the Rashba spin–orbit interaction. This superconducting diode effect enables directional charge transport without energy loss at low temperatures, leading to an ultrahigh sensitivity detection circuit and a modulator with ultralow power consumption, as opposed to the semiconductor diodes used at present, which have high resistivity and are typically unusable at low temperature. In addition, the performance of the superconducting diode is expected to be easily controlled by tuning the superlattice structure, such as the constituent elements, thickness or repetition number. This superconducting diode should pave the way towards the development of superconducting devices.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2590-4.

1. Rikken, G. L. J. A. & Raupach, E. Observation of magneto-chiral dichroism. *Nature* **390**, 493–494 (1997).
2. Braun, F. Ueber die Stromleitung durch Schwefelmetalls. *Ann. Phys.* **153**, 556–563 (1874).
3. Rikken, G. L. J. A. & Wyder, P. Electrical magnetochiral anisotropy. *Phys. Rev. Lett.* **87**, 236602 (2001).
4. Rikken, G. L. J. A. Magnetoelectric anisotropy in diffusive transport. *Phys. Rev. Lett.* **94**, 016601 (2005).
5. Pop, F., Auban-senzier, P., Canadell, E., Rikken, G. L. J. A. & Avarvari, N. Electrical magnetochiral anisotropy in a bulk chiral molecular conductor. *Nat. Commun.* **5**, 3757 (2014).
6. Morimoto, T. & Nagaosa, N. Chiral anomaly and giant magnetochiral anisotropy in noncentrosymmetric Weyl semimetals. *Phys. Rev. Lett.* **117**, 146603 (2016).
7. Ideue, T. et al. Bulk rectification effect in a polar semiconductor. *Nat. Phys.* **13**, 578–583 (2017).
8. Tokura, Y. & Nagaosa, N. Nonreciprocal responses from noncentrosymmetric quantum materials. *Nat. Commun.* **9**, 3740 (2018).
9. Choe, D. et al. Gate-tunable giant nonreciprocal charge transport in noncentrosymmetric oxide interfaces. *Nat. Commun.* **10**, 4510 (2019).
10. Wakatsuki, R. et al. Nonreciprocal charge transport in noncentrosymmetric superconductors. *Sci. Adv.* **3**, e1602390 (2017).
11. Qin, F. et al. Superconductivity in a chiral nanotube. *Nat. Commun.* **8**, 14465 (2017).
12. Yasuda, K. et al. Nonreciprocal charge transport at topological insulator/superconductor interface. *Nat. Commun.* **10**, 2734 (2019).
13. Hoshino, S., Wakatsuki, R., Hamamoto, K. & Nagaosa, N. Nonreciprocal charge transport in two-dimensional noncentrosymmetric superconductors. *Phys. Rev. B* **98**, 054510 (2018).
14. Bychkov, Y. A. & Rashba, I. E. Properties of a 2D electron gas with lifted spectral degeneracy. *JETP Lett.* **39**, 78–81 (1984).
15. LaShell, S., Mcdougall, B. A. & Jensen, E. Spin splitting of an Au (111) surface state band observed with angle resolved photoelectron spectroscopy. *Phys. Rev. Lett.* **77**, 3419–3422 (1996).
16. Ishizaka, K. et al. Giant Rashba-type spin splitting in bulk BiTeI. *Nat. Mater.* **10**, 521–526 (2011).
17. Lustikova, J. et al. Vortex rectenna powered by environmental fluctuations. *Nat. Commun.* **9**, 4922 (2018).
18. Pradipto, A. et al. Enhanced perpendicular magnetocrystalline anisotropy energy in an artificial magnetic material with bulk spin-momentum coupling. *Phys. Rev. B* **99**, 180410 (2019).
19. Nishimura, T. et al. Fabrication of ferrimagnetic Co/Gd/Pt multilayers with structural inversion symmetry breaking. *J. Magn. Soc. Jpn* **44**, 9–14 (2020).
20. Ando, F. et al. Fabrication of noncentrosymmetric Nb/V/Ta superlattice and its superconductivity. *J. Magn. Soc. Jpn* **43**, 17–20 (2019).
21. Wakatsuki, R. & Nagaosa, N. Nonreciprocal current in noncentrosymmetric Rashba superconductors. *Phys. Rev. Lett.* **121**, 026601 (2018).
22. Yip, S. Noncentrosymmetric superconductors. *Annu. Rev. Condens. Matter Phys.* **5**, 15–33 (2014).
23. Bauer, E. & Sigrist, M. *Non-Centrosymmetric Superconductors: Introduction and Overview* (Springer, 2012).
24. Gor'kov, L. P. & Rashba, E. I. Superconducting 2D system with lifted spin degeneracy: mixed singlet-triplet state. *Phys. Rev. Lett.* **87**, 037004 (2001).
25. Frigeri, P., Agterberg, D. F., Koga, A. & Sigrist, M. Superconductivity without inversion symmetry: MnSi versus CePt$_3$Si. *Phys. Rev. Lett.* **92**, 097001 (2004).
26. Fujimoto, S. Electron correlation and pairing states in superconductors without inversion symmetry. *J. Phys. Soc. Jpn* **76**, 051008 (2007).
27. Yanase, Y. & Sigrist, M. Superconductivity and magnetism in non-centrosymmetric system: application to CePt$_3$Si. *J. Phys. Soc. Jpn* **77**, 124711 (2008).
28. Smidman, M., Salamon, M. B., Yuan, H. Q. & Agterberg, D. F. Superconductivity and spin-orbit coupling in non-centrosymmetric materials: a review. *Rep. Prog. Phys.* **80**, 036501 (2017).
29. Edelstein, V. M. Characteristics of the Cooper pairing in two-dimensional noncentrosymmetric electron systems. [*Sov. Phys. JETP* **68**, 1244 (1989)] *Zh. Eksp. Teor. Fiz.* **95**, 2151 (1989).
30. Itahashi, Y. M. et al. Nonreciprocal transport in gate-induced polar superconductor SrTiO$_3$. *Sci. Adv.* **6**, eaay9120 (2020).

Methods

Device fabrication

The multilayer of [Nb (1.0 nm)/V (1.0 nm)/Ta (1.0 nm)]$_{40}$/SiO$_2$ (5.0 nm) was deposited on a MgO (100) substrate using a d.c. magnetron sputtering method[20]. The deposition rates were 0.35 Å s^{-1}, 0.21 Å s^{-1} and 0.44 Å s^{-1} for the Nb, V and Ta targets, respectively, and the MgO substrate was heated at 973 K during the deposition of the [Nb/V/Ta]$_n$ superlattice. To prevent oxidation, a 5.0-nm-thick SiO$_2$ layer was stacked onto the superlattice at room temperature (around 300 K). Next, the deposited film was patterned onto a 50-µm-wide wire using conventional photolithography and an Ar ion milling process. Finally, for the current injection, a Ti (5.0 nm)/Au (100 nm) metal electrode was deposited on the wire. To make an Ohmic contact, the SiO$_2$ capping layer was removed by weak Ar ion milling before the electrode deposition.

Direct- and alternating-current transport measurements

A Yokogawa 7651 was used to inject a d.c. current into the wire and the longitudinal d.c. voltage was measured with a Keithley 2182A Nanovoltmeter. For the a.c. current injection, a Keithley 6221 AC and DC Current Source was used and both the first- and second-harmonic signals were measured with a LI 5640 (NF Corporation).

Details of band structure calculation

To identify the Rashba spin–orbit interaction in the artificial superlattice [Nb/V/Ta]$_n$ without a centre of inversion, we carried out density functional theory calculations for the slab [Nb/V/Ta]$_n$ using the full-potential linearized augmented plane wave+local orbitals method within the generalized gradient approximation in the WIEN2k package[31,32]. We created a slab [Nb/V/Ta]$_5$ containing 30 atoms, which corresponds to stacking two layers of Nb, V and Ta of a body-centred cubic structure five times. We used 31 × 31 × 1 k-point sampling for the self-consistent calculation and the muffin-tin radius R_{MT} of 2.50 atomic units for all atoms. The plane-wave cutoff was given by $R_{\mathrm{MT}}K_{\max} = 8.0$, where $K_{\max}$ is the maximum reciprocal lattice vector. Extended Data Fig. 1 shows the band structure along the high-symmetry line. We obtained the Rashba splitting at the Fermi level near the M points. The magnitude of the Rashba splitting is around 10 meV, which may originate from the V atoms. We thus verified that the [Nb/V/Ta]$_n$ superlattice is a Rashba superconductor.

Nonreciprocal component ΔI_c of a Nb control sample

As a control experiment, we prepared a 120-nm-thick Nb film and carried out the same d.c. measurement. The multilayer Nb (120 nm)/SiO$_2$ (5.0 nm) was deposited on a MgO (100) substrate at 973 K by d.c. magnetron sputtering and processed onto a 50-µm-wide wire structure. As shown in the inset of Extended Data Fig. 2, the T_c of the Nb film is 9.2 K. We then measured the R–I curves to obtain ΔI_c plots at 8.0 K. Although the field dependence of ΔI_c was clearly observed in the temperature range 3.0–4.3 K for the [Nb/V/Ta]$_n$ superlattice ($T_c = 4.41$ K), no difference in ΔI_c was observed for the Nb control sample when changing the field direction. Therefore, the superconducting diode effect can be attributed to the asymmetric structure of the [Nb/V/Ta]$_n$ superlattice.

Coherence length of [Nb/V/Ta]$_n$ superlattice

We investigated the coherence length of the [Nb/V/Ta]$_n$ superlattice to check whether the vortices exist or not when the superconducting diode effect is observed. The coherence length ξ is equivalent to the size of the vortex core around which the supercurrent circulates in type-II superconductors. Through an emergence of the vortex and the increase of the kinetic energy of the supercurrent, a magnetic field destroys Cooper pairs in type-II superconductors. The orbital limiting field B_{c2}^{orb}, referring to the critical field at which vortex cores begin to overlap, is given as $B_{c2}^{\mathrm{orb}} = \Phi_0/2\pi\xi^2$, where $\Phi_0 = 2.07 \times 10^{-15}$ T m^2 is the flux quantum. Here, B_{c2}^{orb} is usually calculated from the initial slope of the plot of B_{c2} versus temperature around T_c by using the Werthamer–Helfand–Hohenberg formula as $B_{c2}^{\mathrm{orb}}(0) = -0.69T_c(\mathrm{d}B_{c2}/\mathrm{d}T)_{T_c}$ in the dirty limit[33]. Therefore, we can estimate the coherence length ξ from the temperature dependence of B_{c2} at T_c. As shown in Extended Data Fig. 3, we obtained the first-harmonic sheet resistances R_ω as a function of magnetic field at $T = 4.0$–4.35 K. The measurement setup and procedure were the same as those shown in the main text. Here, the middle point of the R_ω curve during the transition is defined as the B_{c2}, and the temperature dependence of B_{c2} is plotted in the inset of Extended Data Fig. 3. As a result of linear fitting, the orbital limiting field and the coherence length are estimated to be $B_{c2}^{\mathrm{orb}} = 1.9$ T and $\xi = 13$ nm. Thus, the vortices probably penetrate into the [Nb/V/Ta]$_n$ superlattice along the field direction because the coherence length is much shorter than the thickness of the [Nb/V/Ta]$_n$ superlattice (120 nm).

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

31. Blaha, P. et al. *WIEN2k, An Augmented Plane Wave+Local Orbitals Program for Calculating Crystal Properties* (Karlheinz Schwarz, 2018).
32. Blaha, P. et al. WIEN2k: An APW+lo program for calculating the properties of solids. *J. Chem. Phys.* **152**, 074101 (2020).
33. Werthamer, N. R., Helfand, E. & Hohenberg, P. C. Temperature and purity dependence of the superconducting critical field, H_{c2}. III. Electron spin and spin-orbit effects. *Phys. Rev.* **147**, 295–302 (1966).

Acknowledgements We thank Y. Kasahara, Y. Matsuda and K. Ishida for discussions about the superconducting properties of the [Nb/V/Ta]$_n$ superlattice. This work was supported partly by JSPS KAKENHI grants (15H05702, 15H05884, 15H05745, 17H04924, 18K19021, 18H04225, 18H01178, 18H05227, 18H01815, 19K21972 and 26103002), by the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University, and by the Collaborative Research Program of the Institute for Chemical Research, Kyoto University.

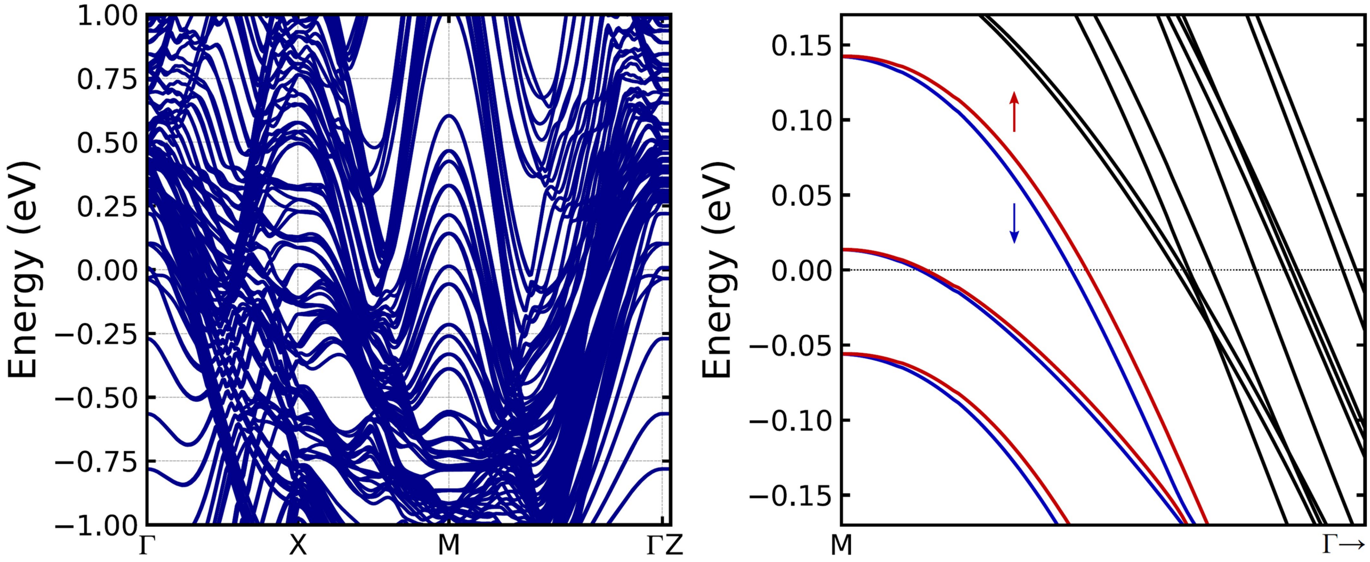
Author contributions T.O. supervised the study. F.A. and Y.M. deposited the films and fabricated them into the devices. F.A. designed the transport measurement setup with help from T.L. and collected the data. T.A. reproduced the experimental results of the superconducting diode effect in another cryogenic equipment. J.I. and Y.Y. calculated the band structure and helped with the analysis of the experimental results. All authors contributed to the interpretation of the results and to the writing of the manuscript.

Competing interests The authors declare no competing interests.

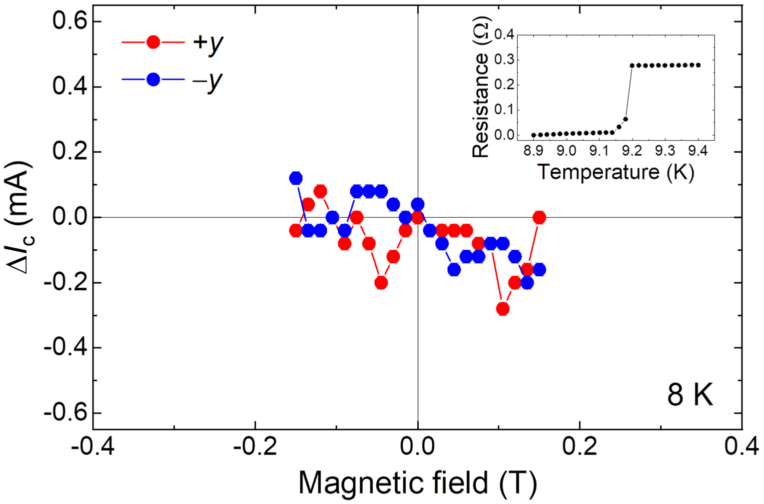
Additional information

Correspondence and requests for materials should be addressed to T.O.

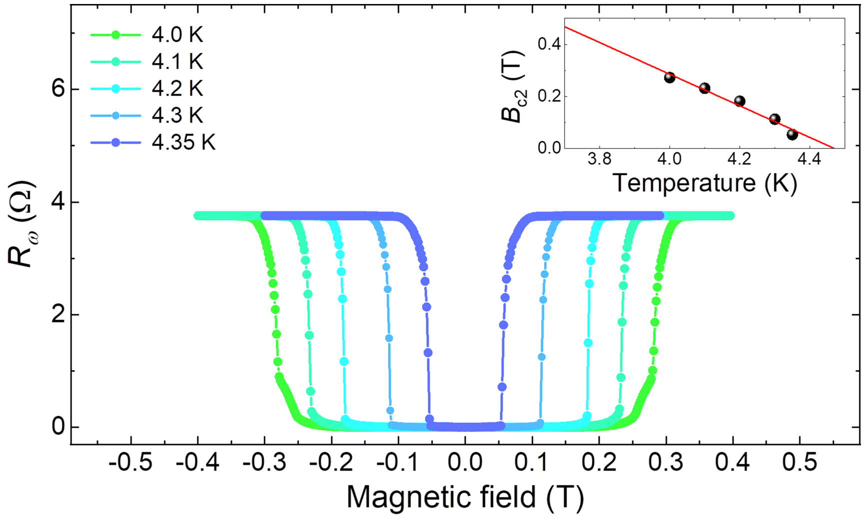
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Band structure of a slab of [Nb/V/Ta]$_5$. **a**, Band structure of a slab [Nb/V/Ta]$_5$ along the high-symmetry line. **b**, Low-energy electron band near the M point.



Extended Data Fig. 2 | The nonreciprocal component of the critical current ΔI_c as a function of magnetic field in a 120-nm-thick Nb film. The inset shows the temperature dependence of the d.c. sheet resistance.



Extended Data Fig. 3 | First-harmonic sheet resistances R_ω of the [Nb/V/Ta]$_n$ superlattice as a function of magnetic field in the vicinity of T_c. The temperature dependence of the critical field B_{c2} is shown in the inset.

Piezoelectric and pyroelectric effects induced by interface polar symmetry

https://doi.org/10.1038/s41586-020-2602-4

Received: 28 January 2020

Accepted: 3 June 2020

Ming-Min Yang[1,5]✉, Zheng-Dong Luo[1], Zhou Mi[2], Jinjin Zhao[2,3], Sharel Pei E[4,6] & Marin Alexe[1]✉

Interfaces in heterostructures have been a key point of interest in condensed-matter physics for decades owing to a plethora of distinctive phenomena—such as rectification[1], the photovoltaic effect[2], the quantum Hall effect[3] and high-temperature superconductivity[4]—and their critical roles in present-day technical devices. However, the symmetry modulation at interfaces and the resultant effects have been largely overlooked. Here we show that a built-in electric field that originates from band bending at heterostructure interfaces induces polar symmetry therein that results in emergent functionalities, including piezoelectricity and pyroelectricity, even though the component materials are centrosymmetric. We study classic interfaces—namely, Schottky junctions—formed by noble metal and centrosymmetric semiconductors, including niobium-doped strontium titanium oxide crystals, niobium-doped titanium dioxide crystals, niobium-doped barium strontium titanium oxide ceramics, and silicon. The built-in electric field in the depletion region induces polar structures in the semiconductors and generates substantial piezoelectric and pyroelectric effects. In particular, the pyroelectric coefficient and figure of merit of the interface are over one order of magnitude larger than those of conventional bulk polar materials. Our study enriches the functionalities of heterostructure interfaces, offering a distinctive approach to realizing energy transduction beyond the conventional limitation imposed by intrinsic symmetry.

Symmetry lies at the heart of the laws of nature that form the basis of modern physics and determine material properties at the fundamental level[5]. Breaking the inversion symmetry allows emergent functionalities and effects. For example, the piezoelectric effect, which converts mechanical energy into electricity and vice versa in a linear manner, is restricted to non-centrosymmetric materials[6]. The pyroelectric effect, which transforms thermal energy into electric energy, occurs only in materials with polar symmetry[7]. Material symmetry is generally determined by its pristine crystallographic structure and loss of symmetry usually occurs via phase transitions. For instance, the paraelectric-to-ferroelectric phase transition in barium titanate ($BaTiO_3$) reduces the symmetry of the crystals from centrosymmetric cubic to polar tetragonal, making $BaTiO_3$ piezoelectric and pyroelectric[6,7]. Nevertheless, the material symmetry can also be tuned by external stimuli that lower the symmetry, or even break the inversion symmetry, of any centrosymmetric material[8,9]. One prominent example is the strain gradient, which parameterizes the inhomogeneity of the strain developed in materials. Strain gradients break the inversion symmetry and induce an electric polarization in materials of any symmetry by the so-called flexoelectric effect[10]. This symmetry breaking is associated with a variety of emergent functionalities, including piezoelectric, pyroelectric and bulk photovoltaic effects, for many materials, such as centrosymmetric strontium titanium oxide ($SrTiO_3$) and titanium dioxide (TiO_2)[11–15]. Despite its universal nature, the real application of this intriguing flexoelectric effect is hampered by its rather small effective coefficients and a complicated setup for inducing large strain gradients. Thus, an alternative would be highly desirable for developing or tuning applications based on induced symmetry breaking.

In this regard, the electric field can play a similar role to the strain gradient in terms of symmetry engineering[8,16]. It has already been employed in two-dimensional systems to engineer their non-centrosymmetry to introduce functionalities with applications in spintronics[17], valleytronics[18] and the photogalvanic effect[19]. The electric field can induce in principle a more general symmetry breaking and not only those mentioned above. As claimed by Nye[8], a crystal under an external stimulus will only show those symmetry elements that are common to both the pristine crystal and the stimulus (Fig. 1a). For example, applying an electric field, which is a vector possessing the conical symmetry of ∞m, to a cubic $SrTiO_3$ crystal with a point symmetry group of $m\bar{3}m$, leads to the common point group of $4mm$, which is polar. Accordingly, the $SrTiO_3$ crystal subjected to the electric field along its (001) direction will no longer show its original cubic symmetry but the polar symmetry (Methods, Extended Data Fig. 1). Therefore, the electric field not only breaks the inversion symmetry but also induces polar structures in centrosymmetric materials. The electric field can be both externally applied and built-in, the latter usually originating from band bending or a chemical potential gradient, which are generally found at heterostructure interfaces. Here we show that

---

[1]Department of Physics, University of Warwick, Coventry, UK. [2]School of Materials Science and Engineering, Shijiazhuang Tiedao University, Shijiazhuang, China. [3]School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang, China. [4]Warwick Manufacturing Group, University of Warwick, Coventry, UK. [5]Present address: Center for Emergent Matter Science, RIKEN, Wako, Japan. [6]Present address: School of Health and Life Sciences, Teesside University, Middlesbrough, UK. ✉e-mail: ymmwhut@gmail.com; m.alexe@warwick.ac.uk

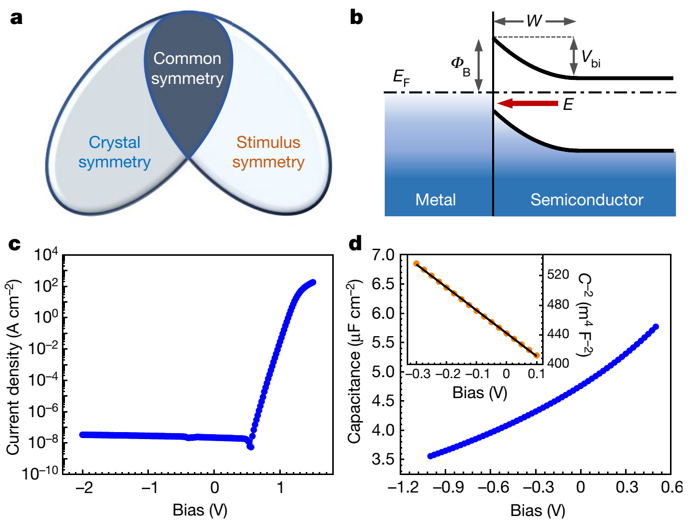


Fig. 1 | Crystal symmetry engineering and Schottky junction electrical characterization. **a**, Schematic of the principle of crystal symmetry engineering by external stimulus. **b**, Schematic of a Schottky junction showing the potential variation in the depletion region, where E_F is the Fermi level, Φ_B is the barrier height, V_{bi} is the built-in potential, W is the depletion region and E denotes the electric field. **c**, **d**, Current–voltage curve (**c**) and capacitance–voltage curve (**d**) of the Au/Nb:SrTiO$_3$/Al junction. The inset in **d** shows the C^{-2} as a function of applied voltage and its linear fit.

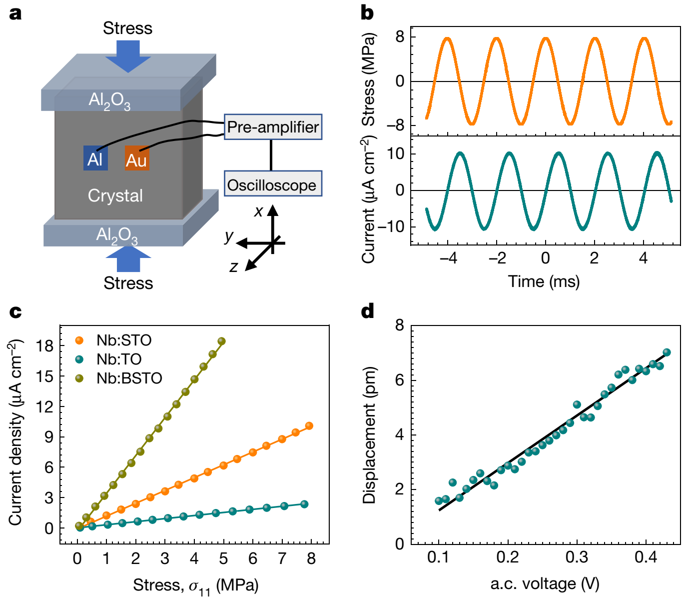


Fig. 2 | Interface piezoelectric effect. **a**, Schematic showing the device used to characterize the direct piezoelectric effect of Schottky junctions. **b**, Waveform of the current density generated by the Au/Nb:SrTiO$_3$/Al junction under the stimulus of sinusoidally varied stress. **c**, Stress-dependent current density generated in the Au/Nb:SrTiO$_3$/Al, Au/Nb:TiO$_2$/Al and Au/Nb:Bb$_{0.6}$Sr$_{0.4}$TiO$_3$/Al junctions. The solid lines are their linear fits. **d**, Surface displacement of the Au/Nb:SrTiO$_3$ junction as a function of the amplitude of applied a.c. voltage. The line is the linear fit.

an electric field manifesting at interfaces can induce polar symmetry that results in emergent piezoelectric and pyroelectric effects in centrosymmetric materials that are otherwise forbidden. We also show that these interface effects can be not only artificially induced in any heterostructures but also rationally tuned to a magnitude much larger than that of conventional bulk materials.

The model systems that usually show a rather strong built-in field are metal–semiconductor contacts termed Schottky junctions. Rearrangement of the energy levels to align the Fermi level in both the metal and the semiconductor generates band bending and a depletion region within the semiconductor associated with an electric field pointing from the semiconductor to the noble metal (Fig. 1b)[1]. Accordingly, polar structures are induced in the depletion region of the centrosymmetric semiconductors. The coefficient of the induced piezoelectric effect associated with a Schottky junction can be predicted as (Methods)

$$d_{ijk} = Q_{jki3}\chi_3\sqrt{2qN_d\chi_3V_{bi}}, \tag{1}$$

where Q_{jki3} is the electrostriction coefficient, χ_3 is the dielectric permittivity in the field direction, q is the elemental charge, N_d is the effective donor density, V_{bi} is the built-in potential in the Schottky junction and the subscripts of the tensor Q, that is, i, j, k, are the elements of $\{1, 2, 3\}$. For the sake of simplicity, the most basic Schottky model is used here to describe the potential profile at the metal–semiconductor interface without considering, for example, interface insulating layer and interface states[20]. Clearly, the piezoelectric coefficient is determined by the centrosymmetric semiconductor properties, such as the dielectric permittivity and the dopant density. A phenomenological theory has also been established to unravel the microscopic mechanism of the interface piezoelectric effect (Methods, Extended Data Fig. 2). Both direct and converse interface piezoelectric effects arise from the combination of the built-in field and the electrostriction effect.

To quantitatively evaluate the piezoelectric coefficient, high-quality Schottky junctions have been fabricated by sputtering noble metal (that is, gold) on (001)-oriented niobium (Nb)-doped SrTiO$_3$ (Nb:STO) and Nb-doped TiO$_2$ (Nb:TO) single crystals (Methods). For the Au/Nb:STO junction, generic electrical properties have been determined by performing current–voltage and capacitance–voltage measurements (Fig. 1c, d). Note that aluminium evaporated on the same surface of the Nb:STO crystal forms Ohmic contacts, which are used as the counter-electrodes with the Schottky junctions (Extended Data Fig. 3). The Au/Nb:STO junction shows an excellent rectification effect with a current density ratio reaching about 10^9 at ±1.5 V and a large capacitance at zero external bias ($C = 4.7$ μF cm^{-2}). The dependence of the reciprocal value of squared capacitance on the external bias is given by[1]

$$C(V)^{-2} = \frac{2V_{bi}}{q\chi_3N_d} - \frac{2V}{q\chi_3N_d}. \tag{2}$$

By performing linear fitting of $C(V)^{-2}$ versus the external applied bias (V), we obtain the values for following parameters: $\chi_3 = 1.68\times10^{-9}$ C V^{-1} m^{-1} (relative dielectric constant $\varepsilon_r = 190$) and $V_{bi} = 1.43$ V (inset of Fig. 1d). From the Hall effect, we obtain the doping density $N_d = 2.4\times10^{25}$ m^{-3}. Given the Nb:STO electrostriction coefficient $Q_{11} = 0.046$ m^4 C^{-2} and $Q_{12} = -0.013$ m^4 C^{-2} (ref. [21]), the corresponding piezoelectric coefficients are estimated from equation (1) to be $d_{33} = 10$ pm V^{-1} and $d_{31} = -3$ pm V^{-1}. These coefficients are of the same order of magnitude of widely used piezoelectric materials such as lithium niobate (LiNbO$_3$; $d_{31} = -2.59$ pm V^{-1})[22].

To experimentally verify the existence and quantitatively evaluate the magnitude of the interface piezoelectric effect in Schottky junctions, we measured the direct piezoelectric effect by applying a dynamic stress to the parallel crystal edges and measuring the short-circuit current generated by the junction (Fig. 2a, Methods). Particular care has been taken to apply the stress homogeneously, minimizing any contributions from the inhomogeneous strain and thus the flexoelectric effect[10]. As shown in Fig. 2b, under the stimulus of a sinusoidal stress with an amplitude of $\sigma_1 = 7.9$ MPa and a frequency of $f = 500$ Hz, the Au/Nb:STO junction outputs an alternative current with the same frequency and an amplitude of $J = 10.1$ μA cm^{-2}. More importantly,

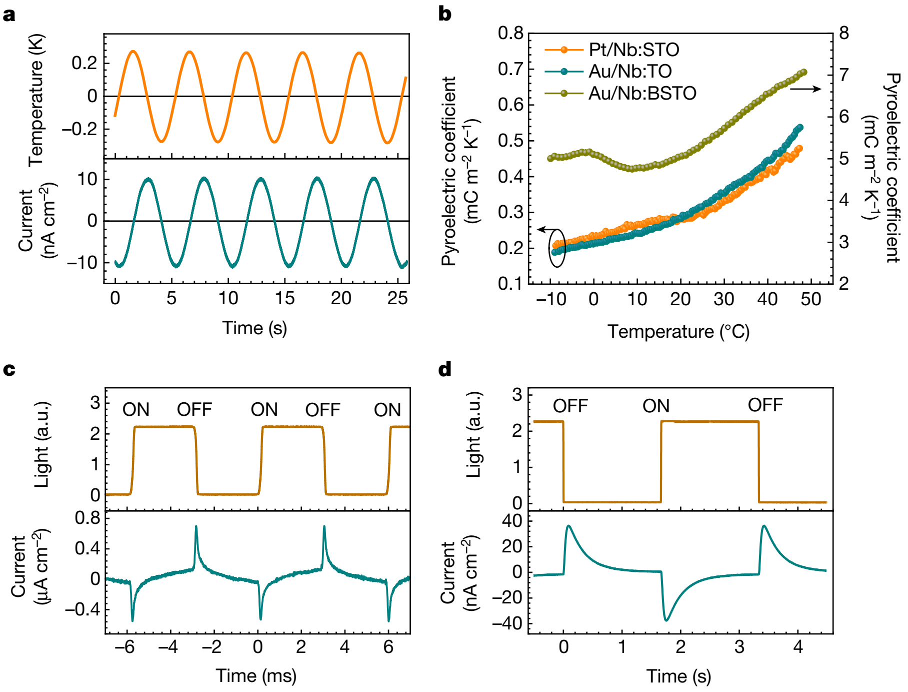


Fig. 3 | Interface pyroelectric effect. a, Waveform of the temperature variation in the Au/Nb:SrTiO$_3$ junction along with the waveform of the generated pyroelectric current density. **b**, Temperature dependence of pyroelectric coefficients of the Au/Nb:SrTiO$_3$, Au/Nb:TiO$_2$ and Au/Nb:BSTO Schottky junctions. **c**, **d**, Pulsed-light-induced transient pyroelectric current in the Au/Nb:BSTO junction (**c**) and the Pb(Ti$_{0.8}$Zr$_{0.2}$)O$_3$ ceramic (**d**).

the amplitude of the output current density increases linearly with the amplitude of the applied stress, demonstrating the manifestation of the direct piezoelectric effect in the Au/Nb:STO junction (Fig. 2c). The corresponding piezoelectric coefficient calculated as $d_{31} = \frac{J}{2\pi f \sigma_1} = -4.07$ pC N^{-1} is close to the value predicted above. To demonstrate that the interface piezoelectricity is a universal effect rather than a phenomenon just limited to the Nb:STO crystals, we performed the same measurements on another centrosymmetric semiconductor, that is, Nb:TiO$_2$ and its Schottky junction with gold. Estimation assuming the same electrostriction coefficient as the SrTiO$_3$ crystal predicts a piezoelectric constant with a magnitude of 1.52 pC N^{-1} for Au/Nb:TO junctions (Extended Data Fig. 3). The measured piezoelectric coefficient of the Au/Nb:TO junctions is about 0.97 pC N^{-1}, which is close to our prediction (Fig. 2c). Note that the Nb:STO and Nb:TO crystals with Ohmic contacts do not show any piezoelectric effect and generate no electricity under the mechanical stimuli, confirming the critical role of the Schottky junctions in generating the piezoelectric effect (Extended Data Fig. 4).

For further confirmation, we explored the converse piezoelectric effect in the Schottky junction by applying an alternative bias to the junction and measuring the resulting surface displacement via atomic force microscopy (Methods). The surface displacement of the Au/Nb:STO junction increases linearly with the amplitude of the excitation voltage, leading to a piezoelectric coefficient of $d_{33} = 16.3$ pm V^{-1}, which is similar to the value estimated above (Fig. 2d). These results clearly demonstrate that the heterostructures of centrosymmetric materials with an interface built-in field have both direct and converse piezoelectric effects, just like the conventional bulk non-centrosymmetric materials.

As mentioned previously, the built-in field within the Schottky junction not merely lifts-off the inversion symmetry but also induces local polarization via the polar nature of the field. Thus, in addition to the piezoelectric effect, the Schottky junction also shows the pyroelectric effect that is the fingerprint feature of any polar structure[7]. This interface pyroelectric effect originates from the temperature dependence of the dielectric permittivity, effective dopant density and built-in potential in Schottky junctions (equation (18) in Methods). To demonstrate this scenario, we measured the pyroelectric effect in Schottky junctions by dynamically modulating their temperature and measuring the generated short-circuit current (Methods). When the temperature of the Au/Nb:STO junction is being sinusoidally modulated, the junction outputs an alternating current with a phase shift of 90°, confirming the manifestation of the pyroelectric effect at Schottky junctions (Fig. 3a). The corresponding pyroelectric coefficient of the Au/Nb:STO junction reaches 298 μC m^{-2} K^{-1} at room temperature. The Au/Nb:TO junction also shows the pyroelectric effect with a room-temperature coefficient of 312 μC m^{-2} K^{-1} (Fig. 3b). Both values are comparable to the values for classical pyroelectric materials[7].

Having demonstrated the interface-polar-symmetry-induced piezoelectricity and pyroelectricity in the Schottky junctions, we further explore their potential by enhancing their coefficients. As indicated by equation (1) and discussed in Methods, the magnitude of both interface piezoelectric and pyroelectric effects depends on the doping density, dielectric permittivity and their tunability with respect to stress, electric field and temperature. Thus, Schottky junctions consisting of semiconductors with a large dielectric tunability are expected to show both enhanced piezoelectric and pyroelectric effects. To this end, we chose 0.1 wt% Nb-doped barium strontium titanium oxide (Ba$_{0.6}$Sr$_{0.4}$TiO$_3$; Nb:BSTO) ceramics to form Schottky junctions with gold. It is known that undoped Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramics show a paraelectric-to-ferroelectric transition around −2 °C, giving rise to a substantial dielectric tunability with a dielectric constant of $\varepsilon_r = 5{,}300$ at room temperature[23]. Nevertheless, both Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ and Nb:BSTO are centrosymmetric at room temperature, being in their cubic phase. The general electrical characterization of the Au/Nb:BSTO junction is given in Extended Data Fig. 5 and the preparation details are given in Methods. As demonstrated in Fig. 2c, this junction shows a substantial piezoelectric effect with a coefficient $d_{31} = -12$ pC N^{-1}, which is about three orders of magnitude higher than that of the undoped Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramics[24]. In contrast, Nb:BSTO

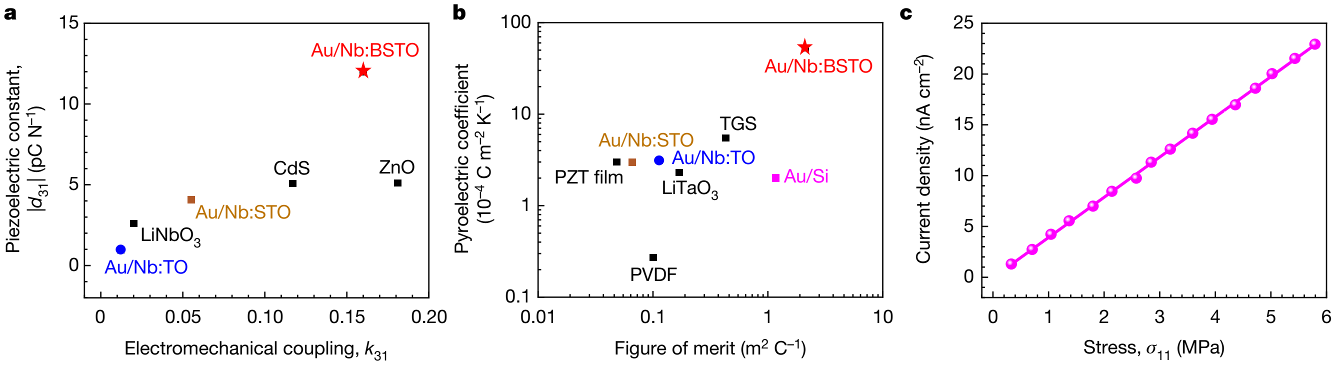


Fig. 4 | Giant magnitude and universal nature of the interface polar effects. **a**, Comparison of the interface piezoelectric constants d_{31} and the electromechanical coupling factors k_{31} of the studied devices with those of conventional polar materials. **b**, Comparison of the pyroelectric coefficients and the figures of merit of the studied devices with those of ferroelectric materials. **c**, The amplitude of the current density generated in Au/Si junctions as a function of the amplitude of the applied stress. Piezoelectric and pyroelectric data on bulk materials are taken from refs. [7,22,25,27]. PZT, lead zirconate titanate ($Pb(Ti_{0.8}Zr_{0.2})O_3$); PVDF, polyvinylidene fluoride; TGS, triglycine sulfate.

ceramics with a quasi-Ohmic contact show a negligible piezoelectric effect (Extended Data Fig. 6). More striking, the Au/Nb:BSTO junction shows a large pyroelectric effect with a room-temperature coefficient reaching 5.3 mC m^{-2} K^{-1} (Fig. 3b). The obtained pyroelectric coefficient here is over one order of magnitude larger than that for conventional ferroelectric materials, such as lithium tantalate ($LiTaO_3$) crystal (230 μC m^{-2} K^{-1}) widely used in the fabrication of pyroelectric detectors[7]. In addition to their large coefficients, the interface pyroelectric effect in Schottky junctions shows another two distinctive features compared with conventional bulk materials. First, the pyroelectric effect in conventional ferroelectric materials has a strong temperature dependence, that is, pyroelectric coefficients decay sharply away from the phase transition temperature, inevitably limiting their working temperature in practical devices. In contrast, the pyroelectric coefficient in the Au/Nb:BSTO junction shows a weak temperature dependence across the phase transition region and retains a large magnitude over a wide temperature range, owing to the persistence of the depletion region (Fig. 3b). Similarly, the pyroelectric coefficients in both Au/Nb:STO and Au/Nb:TO junctions increase monotonically with temperature, supporting their wide working temperature range. Second, the interface pyroelectric effect has a rapid response to the thermal perturbation. Figure 3c, d shows the time dependence of the pyroelectric current generated by the Au/Nb:BSTO junction and a commercial lead titanium zirconium oxide ($Pb(Ti_{0.8}Zr_{0.2})O_3$) ceramic under the same red light pulsed illumination. Clearly, the pyroelectric response of the Au/Nb:BSTO ceramic is over one order of magnitude larger than that of the poled $Pb(Ti_{0.8}Zr_{0.2})O_3$ ceramic. Moreover, the thermal time constant is three orders of magnitude shorter (about 300 μs) than that of the bulk $Pb(Ti_{0.8}Zr_{0.2})O_3$ ceramic (300 ms) of similar dimensions.

We emphasis two main features of these effects arising from the interface polar symmetry. First, both piezoelectric and pyroelectric coefficients observed at the metal–semiconductor interface surpass that of conventional polar materials. Although the interface piezoelectric constants are smaller than those of ferroelectric materials with switchable polarizations (for example, $BaTiO_3$ crystals), they still rival that of non-switchable polar materials, such as zinc oxide (ZnO) and cadmium sulfide (CdS) (Fig. 4a). For example, the piezoelectric constant of the Au/Nb:BSTO junction is over two times larger than that of the ZnO crystals, which have a similar electromechanical coupling factor (Methods)[25]. Apart from oxide semiconductors, there is still a large space to enhance the interface piezoelectric coefficient by exploring a wide range of semiconductors with a large electrostriction effect, such as the organic–inorganic halide perovskites wherein the electrostriction coefficient is over three orders of magnitude larger than that of $SrTiO_3$ crystal[26]. Remarkably, the interface pyroelectric effect is much larger than that of conventional polar materials, even the best ferroelectrics. The Schottky junction shows both a substantial pyroelectric coefficient and a large figure of merit $F_V = p_i/c_p\chi_3$, where p_i is the pyroelectric coefficient and c_p is the heat capacity (Fig. 4b, Methods). In particular, the Au/Nb:BSTO interface shows a figure of merit of 2.11 m^2 C^{-1}, which is one order of magnitude larger than that of classic ferroelectric materials, such as $LiTaO_3$ crystal ($F_V = 0.17$ m^2 C^{-1})[7]. This enhanced figure of merit in the Schottky junction originates from the large pyroelectric coefficient and built-in field-depressed dielectric permittivity in the depletion region.

Interface piezoelectric and pyroelectric effects are universal effects applicable to materials of any symmetry. These effects occur in the heterostructures wherever an electric field builds at the interface. It is worth noting that the electric field is ubiquitous at interfaces of dissimilar materials owing to the chemical potential inhomogeneity across interfaces. To validate this scenario, we studied the piezoelectric and pyroelectric effects of Schottky junctions on silicon wafer. The Au/Si (001) junction outputs a dynamic electrical current, the amplitude of which increases linearly with that of the applied stress (Fig. 4c). This corresponds to a low but finite piezoelectric constant of about −0.013 pC N^{-1}. Moreover, the silicon Schottky junction shows a sizable pyroelectric effect with a room temperature coefficient of 200 μC m^{-2} K^{-1} and a figure of merit of $F_V = 1.17$ m^2 C^{-1} (Fig. 4b).

In summary, we have demonstrated interface piezoelectric and pyroelectric effects that not only show substantial coefficients but also are free from the symmetry limitation. They can be found in and are applicable to a wide range of materials, from conventional semiconductors and oxides, to halide perovskites and two-dimensional materials. These features enable their practical applications in the realm of electromechanical and thermal effects, such as energy conversion and infrared sensors, with distinctive mechanisms and additional tuning feasibility that are different from that of intrinsic non-centrosymmetric materials. With careful design, the interface polar effects can also work concurrently with bulk effects arising from inherent[7,8] or externally induced polarity by, for example, strain gradients[10–15], to achieve enhanced piezoelectric and pyroelectric coefficients or even new effects.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2602-4.

1. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* (John Wiley & Sons, 2006).
2. Fahrenbruch, A. & Bube, R. *Fundamentals of Solar Cells: Photovoltaic Solar Energy Conversion* (Academic Press, 1983).
3. Klitzing, K. V., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett*. **45**, 494–497 (1980).
4. Gozar, A. et al. High-temperature interface superconductivity between metallic and insulating copper oxides. *Nature* **455**, 782–785 (2008).
5. Livio, M. Why symmetry matters. *Nature* **490**, 472–473 (2012).
6. Mason, W. P. & Baerwald, H. *Piezoelectric Crystals and their Applications to Ultrasonics* (D. Van Nostrand Company, 1950).
7. Whatmore, R. Pyroelectric devices and materials. *Rep. Prog. Phys*. **49**, 1335–1386 (1986).
8. Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (Oxford Univ. Press, 1985).
9. Bir, G. L. & Pikus, G. E. *Symmetry and Strain-induced Effects in Semiconductors* (John Wiley & Sons, 1974).
10. Zubko, P., Catalan, G. & Tagantsev, A. K. Flexoelectric effect in solids. *Annu. Rev. Mater. Res*. **43**, 387–421 (2013).
11. Zubko, P., Catalan, G., Buckley, A., Welche, P. R. & Scott, J. F. Strain-gradient-induced polarization in $SrTiO_3$ single crystals. *Phys. Rev. Lett*. **99**, 167601 (2007).
12. Narvaez, J., Vasquez-Sancho, F. & Catalan, G. Enhanced flexoelectric-like response in oxide semiconductors. *Nature* **538**, 219–221 (2016).
13. Yang, M.-M., Kim, D. J. & Alexe, M. Flexo-photovoltaic effect. *Science* **360**, 904–907 (2018).
14. Yang, M. M., Iqbal, A. N., Peters, J. J. P., Sanchez, A. M. & Alexe, M. Strain-gradient mediated local conduction in strained bismuth ferrite films. *Nat. Commun*. **10**, 2791 (2019).
15. Meirzadeh, E. et al. Surface pyroelectricity in cubic $SrTiO_3$. *Adv. Mater*. **31**, 1904733 (2019).
16. Cheong, S.-W. SOS: symmetry-operational similarity. *npj Quantum Mater*. **4**, 53 (2019).
17. Papadakis, S., De Poortere, E., Manoharan, H., Shayegan, M. & Winkler, R. J. S. The effect of spin splitting on the metallic behavior of a two-dimensional system. *Science* **283**, 2056–2058 (1999).
18. Wu, S. et al. Electrical tuning of valley magnetic moment through symmetry control in bilayer MoS_2. *Nat. Phys*. **9**, 149–153 (2013).
19. Yuan, H. et al. Generation and electric control of spin-valley-coupled circular photogalvanic current in WSe_2. *Nat. Nanotechnol*. **9**, 851–857 (2014).
20. Suzuki, S. et al. Fabrication and characterization of $Ba_{1-x}K_xBiO_3$/Nb-doped $SrTiO_3$ all-oxide-type Schottky junctions. *J. Appl. Phys*. **81**, 6830–6836 (1997).
21. Kavasov, A. & Tagantsev, A. K. Positive effective Q_{12} electrostrictive coefficient in perovskites. *J. Appl. Phys*. **112**, 094106 (2012).
22. Yamada, T., Niizeki, N. & Toyoda, H. Piezoelectric and elastic properties of lithium niobate single crystals. *Jpn J. Appl. Phys*. **6**, 151–155 (1967).
23. Samara, G. A. Pressure and temperature dependences of the dielectric properties of the perovskites $BaTiO_3$ and $SrTiO_3$. *Phys. Rev*. **151**, 378–386 (1966).
24. Biancoli, A., Fancher, C. M., Jones, J. L. & Damjanovic, D. Breaking of macroscopic centric symmetry in paraelectric phases of ferroelectric materials and implications for flexoelectricity. *Nat. Mater*. **14**, 224–229 (2015).
25. Kobiakov, I. B. Elastic, piezoelectric and dielectric properties of ZnO and CdS single crystals in a wide range of temperatures. *Solid State Commun*. **35**, 305–310 (1980).
26. Chen, B. et al. Large electrostrictive response in lead halide perovskites. *Nat. Mater*. **17**, 1020–1026 (2018); correction **17**, 1164 (2018).
27. Mangalam, R. V. K., Agar, J. C., Damodaran, A. R., Karthik, J. & Martin, L. W. Improved pyroelectric figures of merit in compositionally graded $PbZr_{1-x}Ti_xO_3$ thin films. *ACS Appl. Mater. Interfaces* **5**, 13235–13241 (2013).

Methods

Symmetry analysis of (001)-oriented Nb:SrTiO$_3$ and Nb:TiO$_2$ Schottky junctions

The Nb:SrTiO$_3$ single crystal belongs to the point symmetry group of $m\bar{3}m$, which includes the symmetry elements of $(1, 2_{\{100\}}, 2_{\{110\}}, 3, 4, \bar{1}, \bar{3}, \bar{4}, m_{\{100\}}, m_{\{110\}})$. The electrical field in the Schottky junction of Nb:SrTiO$_3$ crystal points along the (001) direction. Owing to its vector nature, the field shows the symmetry of ∞m, which includes two types of symmetry elements, that is, infinite rotation symmetry along (001) direction and infinite mirror symmetry. The ∞m symmetry can be represented by a cone. Owing to the manifestation of the electrical field in the Schottky junction, the depletion region will only show the point symmetry that is the subgroup to both $m\bar{3}m$ and ∞m. As illustrated in Extended Data Fig. 1a, the symmetry elements common to both symmetry groups are $(1, 2_{(001)}, 4_{(001)}, m_{(100)}, m_{(010)}, m_{(110)}, m_{(1\text{-}10)})$. The resultant group of symmetry elements corresponds to the point group of $4mm$, which represents polar structures, such as that of BaTiO$_3$ in the tetragonal phase. Similarly, the rutile Nb:TiO$_2$ possesses the point group of $4/mmm$, which includes symmetry elements of $(1, 2_{\{100\}}, 2_{(1\text{-}10)}, 2_{(110)}, 4_{(001)}, \bar{1}, \bar{4}, m_{\{100\}}, m_{(1\text{-}10)}, m_{(110)})$. Its common subgroup with ∞m is also the point group $4mm$ (Extended Data Fig. 1b).

Interface piezoelectricity at Schottky junction

If the work function of the metal exceeds that of the n-type semiconductor, a Schottky barrier forms at the interface between the metal and the semiconductor (Fig. 1b). In the ideal case, the depletion width W is given by[1]

$$W = \sqrt{\frac{2\chi_3}{qN_\mathrm{d}}\left(V_\mathrm{bi} - V - \frac{k_\mathrm{B}T}{q}\right)}, \tag{3}$$

where χ_3 is the dielectric permittivity, q is the electron charge, N_d is the density of dopant, V_bi is the built-in voltage, V is the external applied bias, k_B is the Boltzmann constant and T is the absolute temperature. As the term $k_\mathrm{B}T/q$ is usually much smaller than V_bi in the case of interest, equation (3) can be simplified as

$$W = \sqrt{\frac{2\chi_3}{qN_\mathrm{d}}(V_\mathrm{bi} - V)}. \tag{4}$$

The potential variation in the depletion region is given as[1]

$$V(x) = \frac{qN_\mathrm{d}}{\chi_3}\left(Wx - \frac{1}{2}x^2\right) - \Phi_\mathrm{B}, \tag{5}$$

where x is the distance away from the metal–semiconductor interface into the depletion region and Φ_B is the barrier height at the metal–semiconductor interface. Thus, the corresponding electric field is

$$E(x) = \frac{\partial V}{\partial x} = \frac{qN_\mathrm{d}}{\chi_3}(W - x). \tag{6}$$

Therefore, the local strain $\varepsilon(x)$ in the depletion region induced by the electrostriction effect can be predicted as

$$\varepsilon(x) = ME^2 = M\left(\frac{qN_\mathrm{d}}{\chi_3}\right)^2 (W - x)^2, \tag{7}$$

where M is the electrostriction coefficient (in its one-dimensional form) in the unit of m^2 V^{-2}. The total displacement over the depletion region is then

$$\Delta L = \int_0^W \varepsilon(x)\mathrm{d}x = \frac{1}{3}M\left(\frac{qN_\mathrm{d}}{\chi_3}\right)^2 W^3 = \frac{2}{3}M\sqrt{\frac{2qN_\mathrm{d}}{\chi_3}}(V_\mathrm{bi} - V)^{\frac{3}{2}}. \tag{8}$$

If an a.c. voltage with the form

$$V = V_\mathrm{a}\sin(\omega t) + V_0, \tag{9}$$

where V_a is the amplitude, V_0 is the voltage offset and ω is the angular frequency, is applied on the junction, the corresponding displacement would be given by

$$\Delta L = \frac{2}{3}M\sqrt{\frac{2qN_\mathrm{d}}{\chi_3}}[V_\mathrm{bi} - V_\mathrm{a}\sin(\omega t) - V_0]^{\frac{3}{2}}. \tag{10}$$

Owing to the nonlinear exponent, the Schottky junction would generate a first-order harmonic displacement (that is, strain), which can be obtained by calculating the Fourier series of above equation. In the first approximation the displacement is given by

$$\Delta L_{f_0} = MV_\mathrm{a}\sin(\omega t)\sqrt{\frac{2qN_\mathrm{d}}{\chi_3}(V_\mathrm{bi} - V_0)}. \tag{11}$$

Therefore, the Schottky junction behaves similar to a classical piezoelectric material whose strain varies linearly with applied bias. The effective piezoelectric constant d_eff is

$$d_\mathrm{eff} = \frac{\Delta L_{f_0}}{V_\mathrm{a}} = M\sqrt{\frac{2qN_\mathrm{d}}{\chi_3}(V_\mathrm{bi} - V_0)}. \tag{12}$$

By substituting the electrostriction coefficient from M (m^2 V^{-2}) with a more fundamental parameter Q ($M = Q\chi_3^2$) with units of m^4 C^{-2},

$$d_\mathrm{eff} = Q\chi_3\sqrt{2qN_\mathrm{d}\chi_3(V_\mathrm{bi} - V_0)}. \tag{13}$$

According the developed phenomenological theory (see below), the Schottky junction would possess simultaneously direct and converse piezoelectricity with the same coefficient. Thus, the piezoelectric coefficient of the Schottky junction can be given in the tensor form

$$d_{ijk} = Q_{jki3}\chi_3\sqrt{2qN_\mathrm{d}\chi_3(V_\mathrm{bi} - V_0)}. \tag{14}$$

In the case without any external bias, the Schottky junction shows a piezoelectric tensor as

$$d_{ijk} = Q_{jki3}\chi_3\sqrt{2qN_\mathrm{d}\chi_3 V_\mathrm{bi}}. \tag{15}$$

The depletion region in the Schottky junction behaves like an insulating polar thin layer with electric polarization pointing from the semiconductor bulk to the noble metal interface, as indicated by the red arrow in the Extended Data Fig. 2a. In the equilibrium state, this positive end of electric dipole is compensated by the electrons in the metal interface, while the negative charge of the dipole is compensated by the positive charge in depletion region of the semiconductor. As demonstrated in the phenomenology theory, the interface piezoelectric effect originates from the combination of the built-in electric field and the electrostriction effect. Note that, the electrostriction effect not only describes the electric-field-induced strain with a quadratic dependence but also is a measure of the dependence of the dielectric permittivity on external stress.

Once the junction is subjected to an external stress, for example, a tensile stress perpendicular to the junction interface, the dielectric permittivity of the semiconductor will increase due to the positive electrostriction coefficient Q_{11}. This increased permittivity will give rise to an enhanced electric polarization in the depletion region, which breaks the screening equilibrium at the interface. Therefore, the increased polarization will redistribute the charge between the metal and the semiconductors to reach a new equilibrium state. As the Schottky barrier prevents the electrons from directly flowing across the interface, the electron will flow through the external circuit, giving rise to a displacive electric current (Extended Data Fig. 2b). Similarly, when the junction is subjected to an external electric field, the built-in potential and field will change, which will modulate the strain state of the depletion region via the electrostriction effect. This electric-field-modulated strain leads to the converse piezoelectric effect.

Overall, the microscopic processes of the interface effects rely on the tunability of the semiconductor parameters, especially the dielectric permittivity, with respect to external stimuli. As a fundamental parameter, the dielectric permittivity influences almost all the properties of Schottky junctions, such as the capacitance, depletion width, built-in field and voltage. Therefore, the modulation of electric polarization by external stimuli, which is intrinsically associated with the interface piezoelectric and pyroelectric effects, is accompanied by the variation of all the other junction properties. They are entangled to the piezoelectric and pyroelectric effects.

It is worth noting that the interface piezoelectric effect demonstrated here is distinctive from the surface piezoelectricity, the mechanism and coefficients of which remain elusive[28]. It is also different from the flexoelectric effect in semiconductive oxides, the physics of which was constructed based on the surface piezoelectricity, one of the contributions to the flexoelectricity[12,28]. The flexoelectric effect works only with a strain gradient, that is, inhomogeneous strain. In contrast, the interface piezoelectric effect is due to the electric-field-induced polar symmetry and electrostriction effect, which works in any strain state, including non-strained or homogenous and inhomogeneous strained systems.

Preliminary theory of interface pyroelectric effect

The space charge Q_{SC} per unit area in the Schottky junction is given as[1]

$$Q_{\mathrm{SC}}=\sqrt{2q\chi_3 N_{\mathrm{d}}\left(V_{\mathrm{bi}}-V-\frac{k_{\mathrm{B}}T}{q}\right)}. \tag{16}$$

In the case without applying external bias, equation (16) can be rewritten as

$$Q_{\mathrm{SC}}=\sqrt{2q\chi_3 N_{\mathrm{d}}V_{\mathrm{bi}}}. \tag{17}$$

This unit area space charge Q_{SC} can be regarded as the effective polarization of the Schottky junction, which is a function of dielectric permittivity χ_3, dopant density N_{d} and built-in potential V_{bi}. Therefore, the pyroelectric coefficient of the Schottky junction is

$$\begin{aligned}p_{\mathrm{i}}&=\frac{dQ_{\mathrm{SC}}(\chi_3,N_{\mathrm{d}},V_{\mathrm{bi}})}{\mathrm{d}T}\\&=\frac{\partial Q_{\mathrm{SC}}}{\partial\chi_3}\frac{\partial\chi_3}{\partial T}+\frac{\partial Q_{\mathrm{SC}}}{\partial N_{\mathrm{d}}}\frac{\partial N_{\mathrm{d}}}{\partial T}+\frac{\partial Q_{\mathrm{SC}}}{\partial V_{\mathrm{bi}}}\frac{\partial V_{\mathrm{bi}}}{\partial T}\\&=\sqrt{\frac{qN_{\mathrm{d}}V_{\mathrm{bi}}}{2\chi_3}}\frac{\partial\chi_3}{\partial T}+\sqrt{\frac{q\chi_3V_{\mathrm{bi}}}{2N_{\mathrm{d}}}}\frac{\partial N_{\mathrm{d}}}{\partial T}+\sqrt{\frac{q\chi_3N_{\mathrm{d}}}{2V_{\mathrm{bi}}}}\frac{\partial V_{\mathrm{bi}}}{\partial T}\\&=\frac{Q_{\mathrm{SC}}}{\sqrt{2}}\left(\frac{1}{\chi_3}\frac{\partial\chi_3}{\partial T}+\frac{1}{N_{\mathrm{d}}}\frac{\partial N_{\mathrm{d}}}{\partial T}+\frac{1}{V_{\mathrm{bi}}}\frac{\partial V_{\mathrm{bi}}}{\partial T}\right).\end{aligned} \tag{18}$$

According to equation (18), the pyroelectric coefficient p_{i} in a Schottky junction can be enhanced by using materials with a large dielectric tunability and temperature-sensitive dopant density. The detailed temperature dependence of the dielectric permittivity, effective dopant density and built-in potential are material specific, and remain to be resolved case by case.

When the Schottky junction absorbs heat and increases its temperature, the electric polarization generally decreases. This requires a charge redistribution from the metal interface to the semiconductor bulk through an external circuit (Extended Data Fig. 2c). Cooling the Schottky junction will reverse this process and current direction. Thus, the junction outputs a displacive electric current when subjected to a thermal perturbation.

Note that the effective dielectric permittivity χ_3 of the junction is much smaller that of pristine crystal and ceramic. The large built-in field in the depletion region depresses the dielectric permittivity due to its dielectric tunability[29]. This electric-field-modulated permittivity in the depletion region leads to two results. First, the temperature dependence of the effective permittivity in the junction is different from that of the insulating undoped BSTO ceramic shown in Extended Data Fig. 5a. Second, the dielectric permittivity of the Au/Nb:BSTO junction is highly correlated with the other temperature-dependent parameters, such as dopant density. For example, owing to the semiconductive nature, the effective dopant density of the Nb:BSTO ceramic is temperature dependent. Changing the semiconductor temperature will modulate the carrier density, which tailors the build-in field and in turn, dielectric permittivity. This contribution might actually be more important in building the effective pyroelectric coefficient than other parameters. Thus, the pyroelectric effect and its coefficient of the Au/Nb:BSTO junction has a different temperature dependence than that of bare insulating BSTO ceramic.

Schottky junction preparation

The (001)-oriented Nb:SrTiO$_3$ and Nb:TiO$_2$ single crystals (SurfaceNet) were first cleaned by acetone, isopropanol and water in an ultrasonic bath. The crystal surface was then cleaned by oxygen plasma for 60 s before sputtering gold electrodes (Cressington sputter coater 208HR). Owing to this optimized preparation technique, the Schottky junctions show negligible hysteresis in the current–voltage characteristics with a very low reverse-bias current, that is, highly insulating in the reverse-biased conditions. This high interface quality enables high repeatability of the observed effects. The Ohmic contacts are formed by evaporating a Pt (40 nm)/Al (10 nm) bilayer on the crystal surface. The silicon crystals with a resistivity of 0.005 Ω cm and a dopant density of 1.2×10^{25} m^{-3} (Okmetic) were cleaned and etched by buffered oxide etcher for 1 min to remove the SiO$_2$ passive layer. Note that the Schottky contact and the Ohmic contact were set at the same sample surface to achieve the same chemical and mechanical condition for both types of contact during the measurements.

Nb-doped Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramic preparation

Undoped and 0.1 wt% Nb-doped Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramics were prepared by the classic solid-state reaction method. Raw chemical powders TiO$_2$ (99.99%, Alfa Aesar), BaCO$_3$ (99.95%, Alfa Aesar), SrCO$_3$ (99.99%, Alfa Aesar) and Nb$_2$O$_3$ (99.9985%, Alfa Aesar) were mixed in 2-propanol and ball milled for 4 h. The mixed powders were calcined at 1,000 °C for 10 h in air. The reacted powder was ground and compressed into pellets, which were sintered in a tube furnace at 1,400 °C for 10 h in air. The obtained ceramic pellets (relative density of about 96%) were cut by a diamond blade saw into a cuboid shape with parallel edges. To fully activate the Nb dopant electrically, the cut pellets were annealed in the forming gas (95% N$_2$ + 5% H$_2$) at 900 °C for 6 h. Then, the two large-area ceramic surfaces were polished by diamond papers (average diamond particle diameter down to 0.5 μm). A carrier density of about 7×10^{24} m^{-3} was measured by the Hall effect.

Electric properties characterization

Current–voltage and capacitance–voltage of the Schottky junctions were characterized using a Keithley 2636B source meter and Keysight E4980A LCR meter, respectively. The capacitance was measured with an a.c. driven voltage of 100 mV at 1 kHz.

Interface direct piezoelectric effect characterization

The direct piezoelectric effect, that is, converting mechanical energy into electrical energy, was measured by a home-built device (Extended Data Fig. 7). The samples with two parallel sides were clamped between a piezoelectric actuator (P-888.51, PI Ceramic) and a micrometre head (number 153-201, Mitutoyo). The dynamic stress that varies sinusoidally with time was generated by the piezoelectric actuator. The current generated in the Schottky junction was detected by a transimpedance amplifier (DLPCA-200, Femto) and then displayed by an oscilloscope (DSO-X 3034A, Agilent Technologies) or analysed by a lock-in amplifier (SR865A, Stanford Research Systems).

The stress σ exerted by the piezoelectric actuator was calibrated by measuring the dynamic strain ε developed in the sample and calculated via its stiffness c, that is, stress $\sigma_{11} = c_{11}\varepsilon_{11}$. The dynamic strain was measured by gluing a strain gauge ($R = 120\ \Omega$, 632-146, RS Ltd) to the sample surface by epoxy. The resistance R of the strain gauge changes once subjected to a strain, that is,

$$\frac{\Delta R}{R} = 2\varepsilon. \tag{19}$$

The resistance variation of the strain gauge was measured with a Wheatstone bridge and a lock-in amplifier, as illustrated in Extended Data Fig. 7. The input voltage to the Wheatstone bridge was set as 1 V. The strain developed in the studied samples is in the order of magnitude of 10^{-5}, resulting in $\Delta R \ll R$. In this case, the correlation between the strain amplitude ε_0 and the lock-in output root mean square (r.m.s) value $V_{\text{r.m.s.}}$ equals

$$\varepsilon_0 = 2.828\ V_{\text{r.m.s.}}. \tag{20}$$

The stiffness c_{11} of Nb:SrTiO$_3$, Nb:TiO$_2$ and Si crystal is 318.1 GPa, 267.4 GPa and 165.7 GPa, respectively[30–32]. The stiffness of the Nb:Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramic is about 165 GPa (ref. [33]).

The electromechanical coupling factor k_{31} of the Schottky junctions are calculated as

$$k_{31} = \frac{d_{31}}{\sqrt{s_{11}\chi_3}}, \tag{21}$$

where s_{11} is the elastic compliance. The s_{11} of Nb:SrTiO$_3$, Nb:TiO$_2$ and Nb:Ba$_{0.6}$Sr$_{0.4}$TiO$_3$ ceramics are 3.3×10^{-12} Pa^{-1}, 6.78×10^{-12} Pa^{-1}, 6.06×10^{-12} Pa^{-1}, respectively[34,35]. The effective dielectric permittivity of the Schottky junctions is calculated by linear fit according to equation (2). As shown by equation (2), the slope of the C^{-2} versus V linear fit is

$$\text{Slope} = -\frac{2}{q\chi_3 N_\text{d}}. \tag{22}$$

The doping density in these semiconductors can be estimated using their carrier density, which can be characterized by the Hall effect. The doping density of Nb:STO, Nb:TO and Nb:BSTO are measured as 2.4×10^{25} m^{-3}, 3.4×10^{25} m^{-3} and 7×10^{24} m^{-3}, respectively. With the values of these parameters, the calculated permittivity of Au/Nb:STO, Au/Nb:TO and Au/Nb:BSTO are 1.68×10^{-9} C Vm^{-1} ($\varepsilon_\text{r} = 190$), 1.02×10^{-9} C Vm^{-1} ($\varepsilon_\text{r} = 115$) and 9.32×10^{-10} C Vm^{-1} ($\varepsilon_\text{r} = 105$), respectively.

Interface converse piezoelectric effect characterization

As illustrated in Extended Data Fig. 8, the interface converse piezoelectric effect of the Schottky junction was characterized by measuring the surface displacement using an atomic force microscopy system (Park XE-100). A sinusoidal-type a.c. voltage with a variable amplitude of V_a and frequency of 22.5 kHz was applied on the noble metal electrode of the Schottky junction via a tungsten probe. The resultant surface displacement due to the converse piezoelectric effect was probed by the atomic force microscope (AFM) tip (PPP-EFM-50, Nanosensors) in contact mode under a loading force of 25 nN. The experiments were carefully designed, that is, by applying a.c. bias to the gold electrode via a probe and using a conductive AFM tip that forms good electrical contact with the gold electrode, to eliminate any electrostatic contribution in the characterization. The dynamic vibration of the AFM tip is sensed by the position-sensitive photodiode in the AFM system. The position-sensitive photodiode outputs a dynamic $|A - B|$ signal, the magnitude of which is proportional to the surface displacement amplitude Δl. The $|A - B|$ signal is analysed by the lock-in amplifier, which outputs an r.m.s. value $V_{\text{r.m.s.}}$ proportional to the amplitude of $|A - B|$ signal with a ratio of 1.414. The dependence of the $|A - B|$ signal on the tip displacement was calibrated by the force–distance curve, which shows a tip sensitivity of about $\eta = 21.4$ mV nm^{-1} (Extended Data Fig. 8b). Therefore, the Schottky surface vibration amplitude Δl can be estimated as

$$\Delta l = \frac{1.414 V_{\text{r.m.s.}}}{\eta}. \tag{23}$$

Thus, the converse piezoelectric constant d_{33} of the Schottky junction is

$$d_{33} = \frac{1.414 V_{\text{r.m.s.}}}{\eta V_\text{a}}. \tag{24}$$

Interface pyroelectric effect characterization

The interface pyroelectric effect of the Schottky junctions was measured by a home-built device, as schematically shown in Extended Data Fig. 9. The sample was attached to a two-stage Peltier cooler; one stage was used for controlling the global temperature and the other for inducing the alternative temperature variation using a signal generator (TTI TGA1241). The current output by the sample was amplified by a transimpedance amplifier (Femto DPLC 200) and then displayed by the oscilloscope or analysed by the lock-in amplifier. The Peltier plate and the sample were mounted in an aluminium box that can be vacuumed by a membrane pump. The temperature of the sample was varied sinusoidally with respect to time as

$$T = T_0 + \Delta T \sin(2\pi f t), \tag{25}$$

where T_0 is the base temperature, ΔT is the temperature variation amplitude and f is the frequency. The pyroelectric coefficient can be calculated as

$$p_\text{i} = \frac{J}{2\pi f \Delta T}, \tag{26}$$

where J is the amplitude of the measured pyroelectric current density.

To characterize the light-induced pyroelectric current, the samples were mounted in vacuum and illuminated by a red laser on the top electrode with a wavelength of 660 nm and light intensity of 200 mW cm^{-2}. The Pb(Zr$_{0.2}$Ti$_{0.8}$)$_3$ and Nb:BSTO ceramics are of equal size in dimension and volume. Based on Fig. 3c, d, we conclude that the overall behaviour of the Schottky junctions is thin film-like rather than bulk-like, supporting the hypothesis that the signal is generated within a skin layer (depletion width) underneath the surface.

The figure of merit (F_V) of the Schottky junctions is calculated as[7]

$$F_\text{V} = \frac{p_\text{i}}{c_\text{p}\chi_3}, \tag{27}$$

where p_i is the pyroelectric coefficient and c_p is the specific heat capacity. The specific heat capacity of Nb:STO, Nb:TO and Nb:BSTO is about 2.7 J cm^{-3} K^{-1} (ref. [36]). The specific heat capacity of silicon is 1.65 J cm^{-3} K^{-1}.

Phenomenological theory of interface piezoelectricity

The volume density of internal energy U of a body subjected to external stresses σ and electric field E can be expressed in the form[8]

$$\mathrm{d}U = \sigma_{ij}\mathrm{d}\varepsilon_{ij} + E_m\mathrm{d}P_m + T\,\mathrm{d}S, \tag{28}$$

where ε is the strain, P is the electric polarization, T is the temperature and S is the volume density of entropy. Note that all the subscripts of the variables, that is, i, j, m and so on, in this section refer to elements of {1,2,3} and Einstein summation convention is used here. Here, we choose (σ, E, S) as the independent variables, with the (ε, P, T) as the dependent variables. Accordingly, we introduce the enthalpy H per unit volume, defined by

$$H = U - \sigma_{ij}\varepsilon_{ij} - E_m P_m. \tag{29}$$

Hence,

$$\mathrm{d}H = -\,\varepsilon_{ij}d\sigma_{ij} - P_m\mathrm{d}E_m + T\,\mathrm{d}S. \tag{30}$$

As can be seen from above equation, the dependent variables can be expressed as

$$\varepsilon_{ij} = -\frac{\partial H}{\partial \sigma_{ij}}$$

$$P_m = -\frac{\partial H}{\partial E_m} \tag{31}$$

$$T = \frac{\partial H}{\partial S}.$$

Since the entropy is a constant in the adiabatic conditions, the dependent variables of interest (ε, E) are a function of stress σ and polarization P

$$\varepsilon_{ij} = \varepsilon(\sigma_{kl}, E_n);\ P_m = P(\sigma_{kl}, E_n). \tag{32}$$

Expanding the above functions to the second order about the position of zero strain and zero electric polarization, we obtain

$$\begin{aligned}\varepsilon_{ij} &= \frac{\partial \varepsilon_{ij}}{\partial \sigma_{kl}}\sigma_{kl} + \frac{\partial \varepsilon_{ij}}{\partial E_n}E_n \\ &\quad + \frac{1}{2!}\left[\frac{\partial^2 \varepsilon_{ij}}{\partial \sigma_{kl}\partial \sigma_{qr}}\sigma_{kl}\sigma_{qr} + 2\frac{\partial^2 \varepsilon_{ij}}{\partial \sigma_{kl}\partial E_n}\sigma_{kl}E_n + \frac{\partial^2 \varepsilon_{ij}}{\partial E_n\partial E_o}E_nE_o\right]\end{aligned} \tag{33}$$

$$\begin{aligned}P_m &= \frac{\partial P_m}{\partial E_n}E_n + \frac{\partial P_m}{\partial \sigma_{kl}}\sigma_{kl} \\ &\quad + \frac{1}{2!}\left[\frac{\partial^2 P_m}{\partial \sigma_{kl}\partial \sigma_{qr}}\sigma_{kl}\sigma_{qr} + 2\frac{\partial^2 P_m}{\partial \sigma_{kl}\partial E_n}\sigma_{kl}E_n + \frac{\partial^2 P_m}{\partial E_n\partial E_o}E_nE_o\right].\end{aligned} \tag{34}$$

The first two differentiation terms in equations (33) and (34) represent the elastic compliance and the dielectric susceptibility, respectively.

$$\frac{\partial \varepsilon_{ij}}{\partial \sigma_{kl}} = s^P_{ijkl} \tag{35}$$

$$\frac{\partial P_m}{\partial E_n} = \chi^\sigma_{mn}. \tag{36}$$

The fourth-order tensor s^P_{ijkl} is the elastic compliance measured at constant electric polarization and the second-order tensor χ^σ_{mn} denotes the dielectric permittivity measured at constant stress. The second first-order differentiation in equations (33) and (34) represents the converse and direct piezoelectric effects in non-centrosymmetric materials.

$$\frac{\partial \varepsilon_{ij}}{\partial E_n} = \frac{\partial P_n}{\partial \sigma_{ij}} = d_{nij}. \tag{37}$$

For simplicity, we assume the piezoelectric constant remains constant under external stress. Thus,

$$\frac{\partial^2 \varepsilon_{ij}}{\partial \sigma_{kl}\partial E_n} = \frac{\partial d_{nij}}{\partial \sigma_{kl}} = 0 \tag{38}$$

$$\frac{\partial^2 P_m}{\partial \sigma_{kl}\partial \sigma_{qr}} = \frac{\partial d_{mkl}}{\partial \sigma_{qr}} = 0. \tag{39}$$

It is also reasonable to assume that the elastic constant of the crystals remains unchanged under external stress, namely, $\frac{\partial^2 \varepsilon_{ij}}{\partial \sigma_{kl}\partial \sigma_{qr}} = 0$, and the dielectric susceptibility remains constant under small external electric field, that is, $\frac{\partial^2 P_m}{\partial E_n\partial E_o} = 0$ (ref. [6]). Also, the other second-order partial derivatives are correlated

$$\frac{\partial^2 \varepsilon_{ij}}{\partial E_n\partial E_o} = -\frac{\partial^3 H}{\partial \sigma_{ij}\partial E_n\partial E_o} = \frac{\partial^2 P_o}{\partial \sigma_{ij}\partial E_n} = 2M_{ijno}. \tag{40}$$

The derived fourth-rank tensor M_{ijno} is the electrostriction coefficient in the unit of m^2 V^{-2}. Therefore, the strain ε_{ij} and electric polarization P_m induced by external stress σ_{kl} and electric field E_n can be written as

$$\varepsilon_{ij} = s^P_{ijkl}\sigma_{kl} + d_{nij}E_n + M_{ijno}E_nE_o \tag{41}$$

$$P_m = \chi^\sigma_{mn}E_n + d_{mkl}\sigma_{kl} + 2M_{klnm}E_n\sigma_{kl}. \tag{42}$$

In the case of materials without inversion symmetry, the external applied electric field induces mechanical strain via both the converse piezoelectric effect and the electrostriction effect. The strain induced by the electrostriction effect in piezoelectric materials is normally much smaller than that induced by the piezoelectric effect and thus is generally ignored. When just applying external stress to the piezoelectric materials without applying an electric field, it would generate an electric polarization only by the direct piezoelectric effect.

In the case of centrosymmetric materials, piezoelectric constants d_{nij} are all zero. Equations (41) and (42) can be rewritten as

$$\varepsilon_{ij} = s^P_{ijkl}\sigma_{kl} + M_{ijno}E_nE_o \tag{43}$$

$$P_m = \chi^\sigma_{mn}E_n + 2M_{klnm}E_n\sigma_{kl}. \tag{44}$$

According to equation (43), the external electric field can induce mechanical strain through only the electrostriction effect. In contrast, homogenous mechanical stress cannot induce electric polarization along in these materials due to the inversion symmetry. However, the second term on the right-hand side of equation (44) indicates that external stress can modulate electric polarization via the electrostriction effect if there is an electric field E_n, which could be either an externally applied field or a built-in space charge field. The effective piezoelectric effect is given as

$$d_{mkl} = 2M_{klnm}E_n = 2M_{klmn}E_n. \tag{45}$$

This can be understood as the electric field, that is, a unidirectional vector, breaking the inversion symmetry in native centrosymmetric materials, inducing electric polarization and giving rise to a piezoelectric effect via the electrostriction. We can unveil the underlying mechanism by rewriting equation (40) as

$$2M_{ijno} = \frac{\partial^2 \varepsilon_{ij}}{\partial E_n \partial E_o} = \frac{\partial^2 P_n}{\partial E_o \partial \sigma_{ij}} = \frac{\partial \chi_{no}^{\sigma}}{\partial \sigma_{ij}}, \qquad (46)$$

keeping in mind that

$$\frac{\partial \varepsilon_{ij}}{\partial E_n} = \frac{\partial}{\partial E_n}\left(-\frac{\partial H}{\partial \sigma_{ij}}\right) = \frac{\partial}{\partial \sigma_{ij}}\left(-\frac{\partial H}{\partial E_n}\right) = \frac{\partial P_n}{\partial \sigma_{ij}}. \qquad (47)$$

According to equation (46), the electrostriction is a measure of the dependence of dielectric permittivity on external stress[37]. This is termed the converse electrostriction effect. Therefore, the external stress would modulate the dielectric permittivity via the electrostriction effect, giving rise to a change in the electric polarization induced by the electric field E_n.

Similarly, the second term on the left-hand side of equation (43) indicates that the electric field E_n can also induce a converse piezoelectric effect in a centrosymmetric material. To derive the corresponding converse piezoelectric coefficient, we extend the Einstein notation in equation (43)

$$\varepsilon_{ij} = s_{ijkl}^{P}\sigma_{kl} + M_{ijnn}E_n^2 + 2M_{ijnq}E_nE_q. \qquad (48)$$

Note that the subscript $n \neq q$ in above equation. For the case of interest here, the electric field exerted on a centrosymmetric material consists of a constant part E_n, which represents the built-in field, and an alternative component ΔE_n due to an externally applied a.c. voltage. Accordingly, equation (48) can be rewritten as

$$\varepsilon_{ij} = s_{ijkl}^{P}\sigma_{kl} + M_{ijnn}E_n^2 + 2M_{ijnn}E_n\Delta E_n + M_{ijnn}\Delta E_n^2. \qquad (49)$$

The first and second terms on the right-hand side of equation (49) represent the static strain induced by the external stress and the built-in electric field in the space-charged region, respectively; the third term represents the first-order harmonic strain induced by the dynamic electric field, that is, the converse piezoelectric effect; the last term is the second-harmonic strain, which refers to the conventional electrostriction strain. Therefore, the converse piezoelectric coefficient is the third term of equation (49)

$$d_{nij} = 2M_{ijnn}E_n. \qquad (50)$$

In the case that the external field E_q is in a different direction with respect to the built-in field E_n, the external-field-induced strain is represented by the third term on the right-hand side of equation (48), that is, $\varepsilon_{ij} = 2M_{ijnq}E_nE_q$. Clearly, the corresponding piezoelectric coefficient is

$$d_{qij} = 2M_{ijnq}E_n = 2M_{ijqn}E_n. \qquad (51)$$

The piezoelectric coefficient expressed in equations (45), (50) and (51) can be transformed into a unified form given as

$$d_{mkl} = 2M_{klmn}E_n. \qquad (52)$$

In summary, both direct and converse piezoelectric effects with the same coefficients can occur in centrosymmetric materials once they are subjected to an electric field E_n.

Data availability

The data that support the findings of this study are available at the University of Warwick open access research repository (http://wrap.warwick.ac.uk/136971) or from the corresponding authors upon request. Source data are provided with this paper.

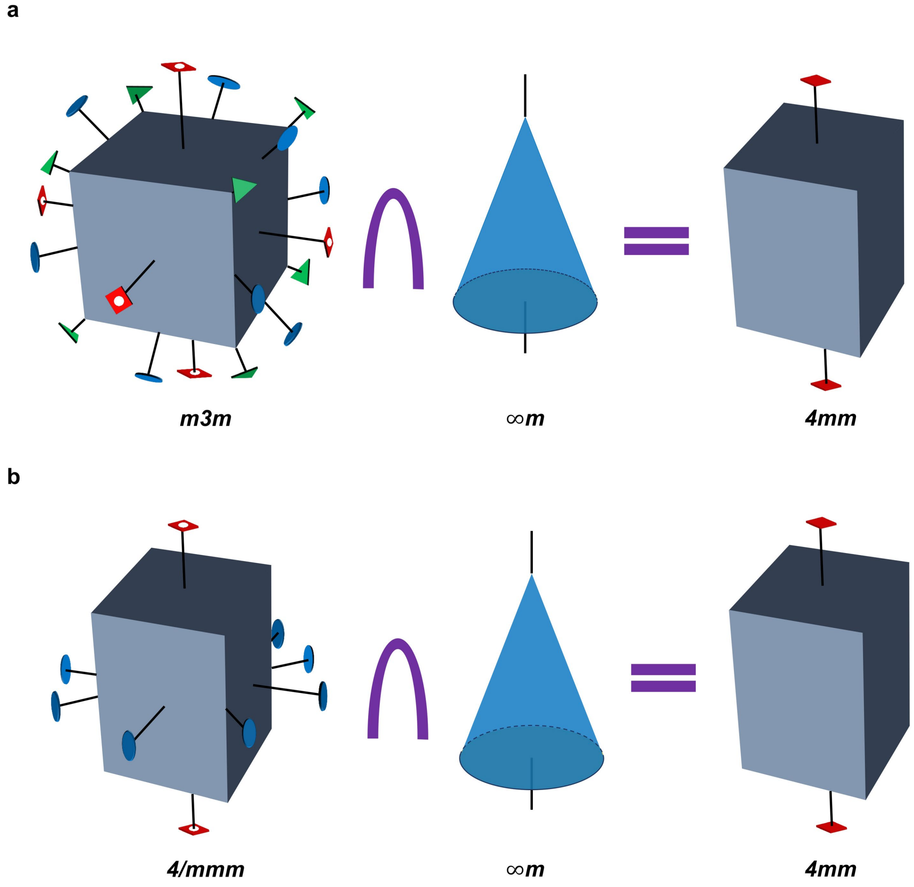
28. Tagantsev, A. K. Piezoelectricity and flexoelectricity in crystalline dielectrics. *Phys. Rev. B* **34**, 5883–5889 (1986).
29. Tagantsev, A. K., Sherman, V. O., Astafiev, K. F., Venkatesh, J. & Setter, N. Ferroelectric materials for microwave tunable applications. *J. Electroceram.* **11**, 5–66 (2003).
30. Schranz, W., Sondergeld, P., Kityk, A. & Salje, E. K. H. Elastic properties of $SrTiO_3$ crystals at ultralow frequencies. *Phase Transit.* **69**, 61–76 (1999).
31. Grimsditch, M. & Ramdas, A. Elastic and elasto-optic constants of rutile from a Brillouin scattering study. *Phys. Rev. B* **14**, 1670–1682 (1976).
32. Mason, W. P. Physical acoustics and the properties of solids. *J. Acoust. Soc. Am.* **28**, 1197–1206 (1956).
33. Kaushal, A., Olhero, S. M., Antunes, P., Ramalho, A. & Ferreira, J. M. F. Structural, mechanical and dielectric properties of $Ba_{0.6}Sr_{0.4}TiO_3$—the benefits of a colloidal processing approach. *Mater. Res. Bull.* **50**, 329–336 (2014).
34. Poindexter, E. & Giardini, A. A. Elastic constant of strontium titanate ($SrTiO_3$). *Phys. Rev.* **110**, 1069 (1958).
35. Wachtman, J. B., Tefft, W. E. & Lam, D. G. Elastic constants of rutile (TiO_2). *J. Res. Natl Bur. Stand. A* **66A**, 465–471 (1962).
36. Durán, A., Morales, F., Fuentes, L. & Siqueiros, J. M. Specific heat anomalies at 37, 105 and 455 K in $SrTiO_3$:Pr. *J. Phys. Condens. Matter* **20**, 085219 (2008).
37. Devonshire, A. F. Theory of ferroelectrics. *Adv. Phys.* **3**, 85–130 (1954).

Acknowledgements M.A. acknowledges the Theo Murphy Blue-sky Awards of Royal Society. The work was partly supported by the EPSRC (UK) through grant numbers EP/M022706/1, EP/P031544/1 and EP/P025803/1. Z.M. and J.Z. acknowledge the National Natural Science Foundation of China (11772207); Natural Science Foundation of Hebei Province for Distinguished Young Scholar (A2019210204) and Shenzhen Peacock Team Program (KQTD20170810160424889). We acknowledge the discussion with H. Zhang, A. N. Iqbal and F. Zhuge; and the technical support from M. Crosbie.

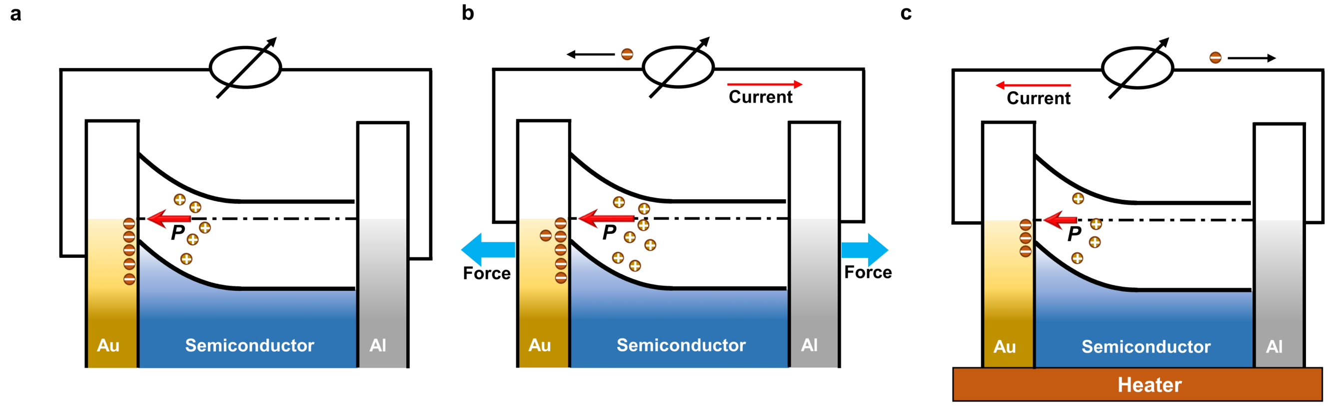
Author contributions M.-M.Y. and M.A. conceived the idea, designed the experiments, collected the data and wrote the manuscript. M.-M.Y. developed the theory. Z.-D.L., Z.M. J.Z. and S.P.E. were involved in sample preparation.

Competing interests The authors declare no competing interests.

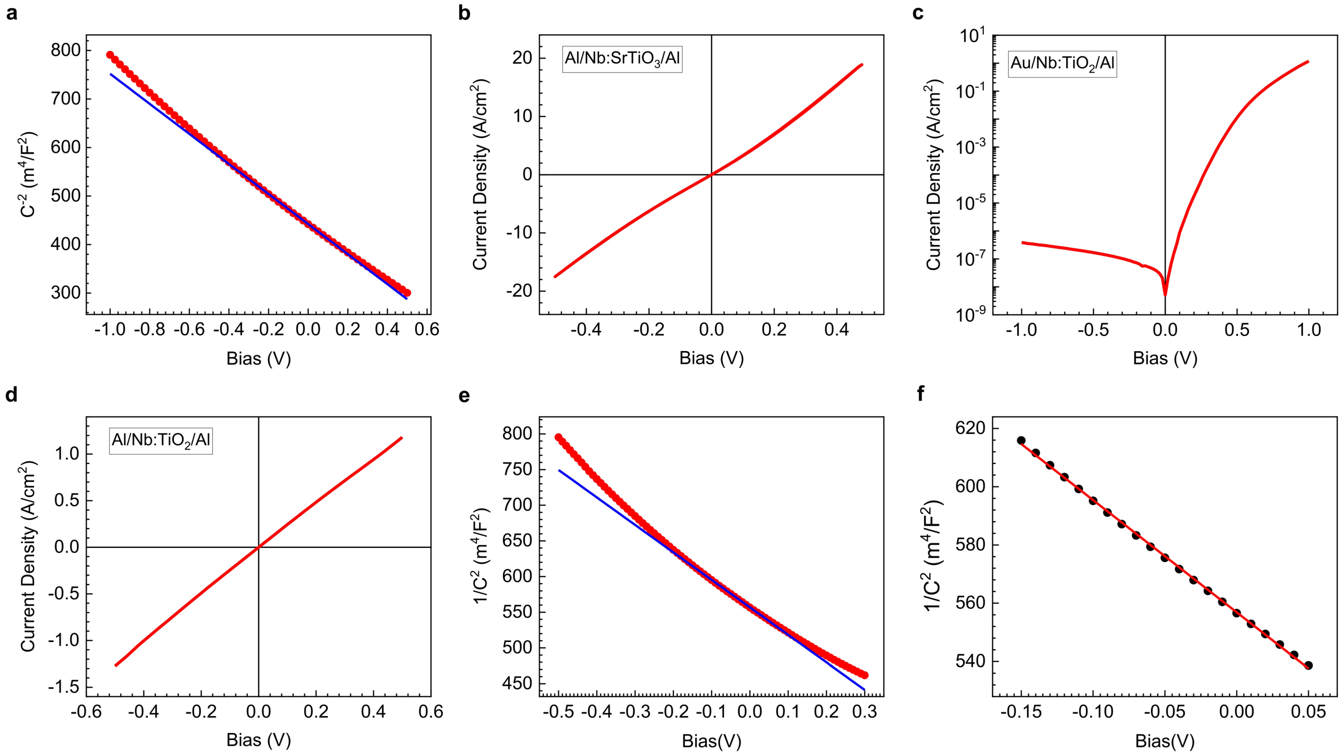
Additional information
Correspondence and requests for materials should be addressed to M.-M.Y. or M.A.
Peer review information *Nature* thanks Long-Qing Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



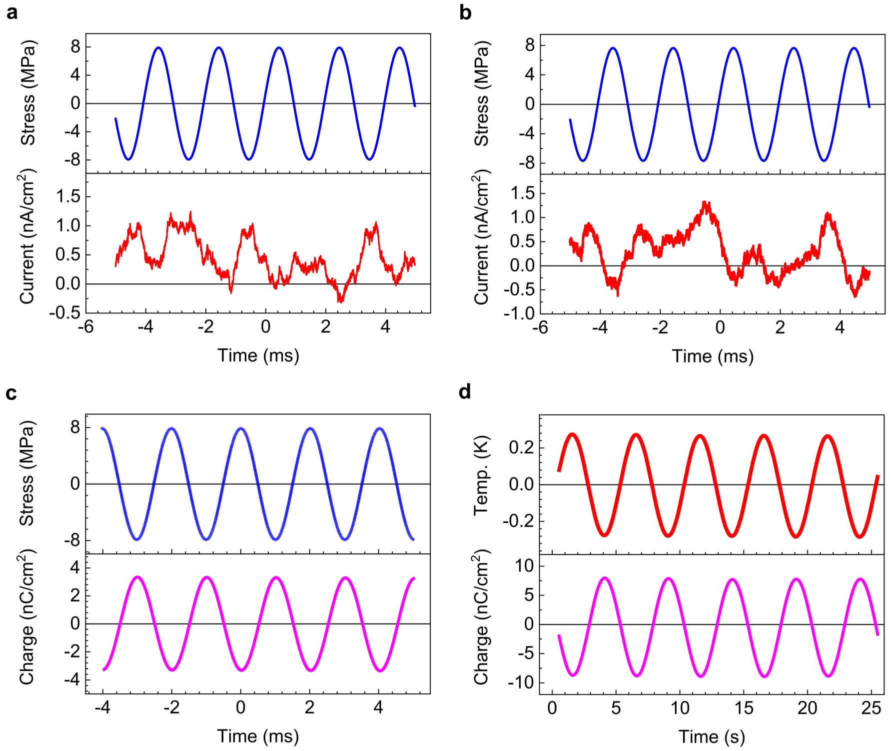
Extended Data Fig. 1 | Electrical-field-engineered symmetry in (001)-oriented Nb:SrTiO$_3$ and Nb:TiO2 crystals. **a**, Schematic showing the common group of the $m\bar{3}m$ point group and the ∞m group. **b**, Schematic showing the common group of the 4/*mmm* point group and the ∞m group. Only the rotation symmetry elements are shown here while the mirror symmetry elements are omitted. The dome symbol represents the intersect operation.



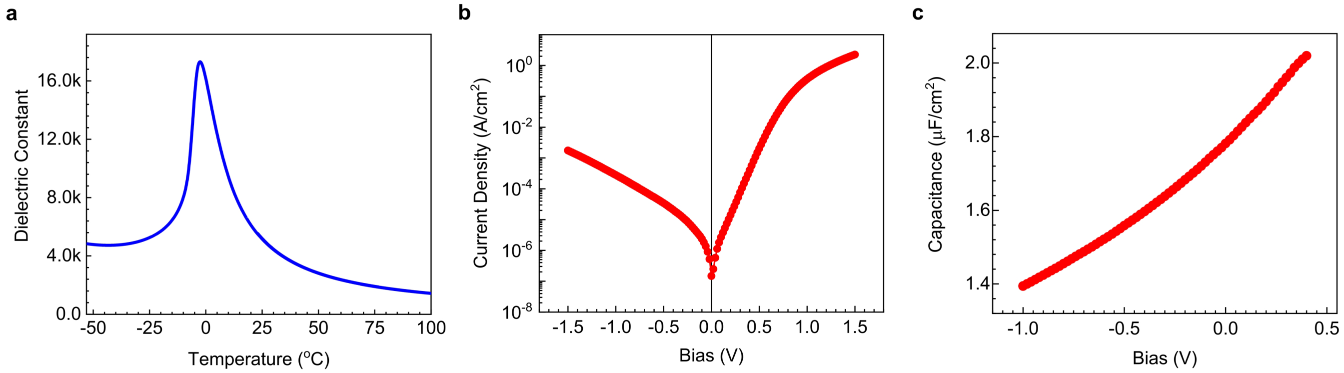
Extended Data Fig. 2 | Microscopic processes of interface piezoelectric and pyroelectric effects. a, The electric polarization and compensating charges of the Schottky junction in the equilibrium state. **b**, Charge redistribution when the junction is subjected to a tensile stress. **c**, The charge redistribution when subjecting to heating. The piezoelectric and pyroelectric effects persist whenever there is a depletion region with a built-in field. However, another factor, that is, the effective barrier, which assures good insulating properties in reverse-bias conditions, is critical for the ability of the junction to deliver displacive current and consequently to output electricity. If the barrier becomes leaky, for example, by further increased temperature, the re-distribution of charge carriers will happen by electron transmission directly cross the interface via either tunnelling or thermionic emission. In this case, the pyroelectric effect might still be there, but it is screened by alternative conducting channels. This is to a certain extent similar to the situation of a solar cell affected by a low shunt resistance.



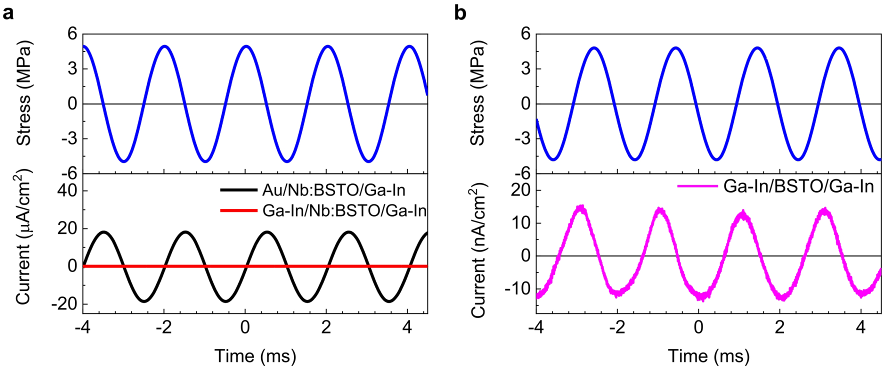
Extended Data Fig. 3 | Electric characterization of the Al/Nb:SrTiO$_3$/Al and Au/Nb:TiO$_2$ junctions. **a**, C^{-2}–V curve of the Au/Nb:SrTiO$_3$ junction in a large voltage range. The red dots are the measured data and the blue line is the linear fit near zero voltage. **b**–**d**, Current–voltage curves of the Al/Nb:SrTiO$_3$/Al (**b**), Au/Nb:TiO$_2$/Al (**c**) and Al/Nb:TiO$_2$/Al (**d**) heterostructures. **e**, C^{-2}–V curve of the Au/Nb:TiO$_2$/Al junction. The red dots are the measured data and the blue line is the linear fit near zero voltage. **f**, The C^{-2}–V curve of the Au/Nb:TiO$_2$/Al junction and its linear fit near zero voltage. Given the dopant density of 3.4×10^{25} m^{-3} in Nb:TiO$_2$, this fit indicates that the effective permittivity of the Au/Nb:TiO$_2$ junction is 1.02×10^{-9} C V m^{-1} and a built-in potential of 1.45 V. As we are mainly concerned here with the piezoelectric effect of the Schottky junctions without applying bias (that is, near-zero voltage), the electrical parameters derived by fitting around the zero-voltage bias give a good description of the junction properties and lead to a quantitative prediction of the piezoelectric effect consistent with experimental results.



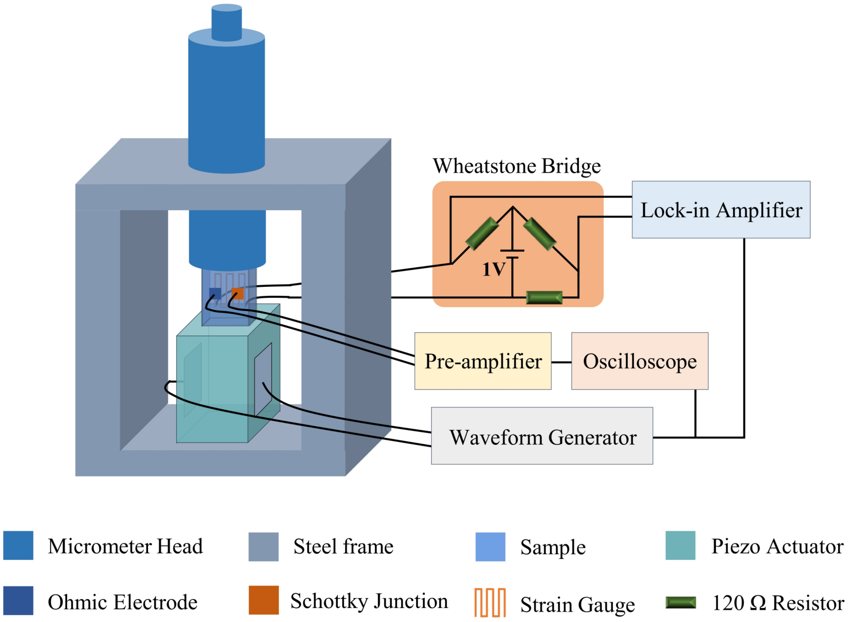
Extended Data Fig. 4 | Current output by the Nb:SrTiO$_3$ and Nb:TiO$_2$ crystals with Ohmic contacts and charges generated in Schottky junctions. **a**, **b**, Current density generated by the the Al/Nb:SrTiO$_3$/Al heterostructure (**a**) and the Al/Nb:TiO$_2$/Al heterostructure (**b**) under the stimuli of external stress. Clearly, the current density waveforms generated in the crystals with Ohmic contacts show not only a low magnitude compared with that shown in Fig. 2 but also an irregular time dependence. This demonstrates both crystals with Ohmic contacts have no piezoelectric effect. **c**, **d**, Charge waveforms generated in Au/Nb:SrTiO$_3$/Al junction driven by dynamic stress (**c**) and temperature (**d**) by integrating the generated current with respect to time in Figs. 2b, 3a.



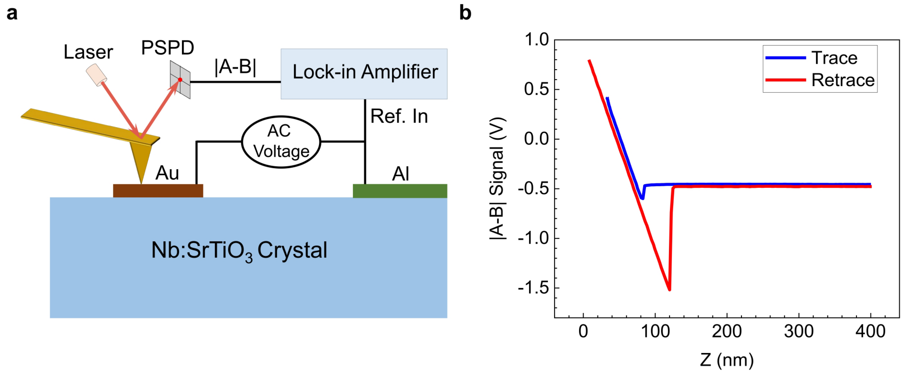
Extended Data Fig. 5 | Electric characterization of the Au/Nb:$Ba_{0.6}Sr_{0.4}TiO_3$ junction. **a**, Temperature-dependent dielectric constant of the insulating $Ba_{0.6}Sr_{0.4}TiO_3$ ceramic. **b**, Current density–voltage curve. **c**, Capacitance–voltage curve of the Au/Nb:$Ba_{0.6}Sr_{0.4}TiO_3$ junction.



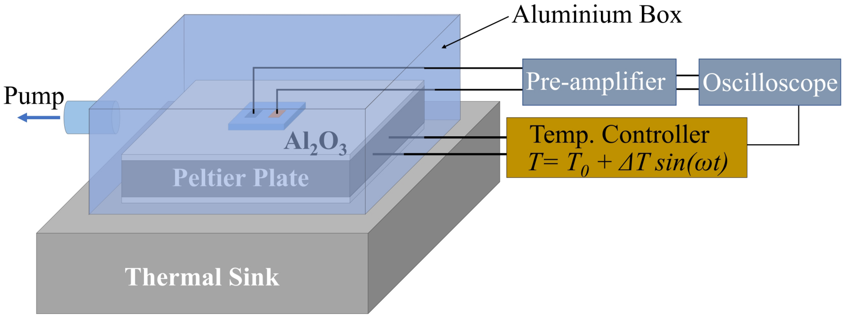
Extended Data Fig. 6 | Negligible piezoelectricity in Ohmic contacted ceramics. **a**, The current density output by the Au/Nb:$Ba_{0.6}Sr_{0.4}TiO_3$/Ga-In (black line) and Ga-In/Nb:$Ba_{0.6}Sr_{0.4}TiO_3$/Ga-In (red line) driven by sinusoidally varied stress (top). **b**, The current density generated by the Ga–In/$Ba_{0.6}Sr_{0.4}TiO_3$/Ga–In heterojunctions under sinusoidally varied stress. Note that the current density amplitude observed in both ceramics with only Ohmic contacts are three to four order of magnitude smaller than that generated in the Au/Nb:$Ba_{0.6}Sr_{0.4}TiO_3$ junction, demonstrating the essential role of the Schottky contact in the induced piezoelectric effect.



Extended Data Fig. 7 | Direct piezoelectric effect characterization setup .



Extended Data Fig. 8 | Converse piezoelectric effect characterization. a, Schematic showing the measurement setup. **b**, Force–distance curve of PPP-EFM-50 (Nanosensors) on the Au/Nb:SrTiO$_3$. *Z* is the distance moved by the AFM tip stage.



Extended Data Fig. 9 | Pyroelectric effect characterization setup .

Liquid–liquid transition and critical point in sulfur

https://doi.org/10.1038/s41586-020-2593-1

Received: 24 April 2019

Accepted: 26 May 2020

Laura Henry[1], Mohamed Mezouar[1]✉, Gaston Garbarino[1], David Sifré[1], Gunnar Weck[2] & Frédéric Datchi[3]

The liquid–liquid transition (LLT), in which a single-component liquid transforms into another one via a first-order phase transition, is an intriguing phenomenon that has changed our perception of the liquid state. LLTs have been predicted from computer simulations of water[1,2], silicon[3], carbon dioxide[4], carbon[5], hydrogen[6] and nitrogen[7]. Experimental evidence has been found mostly in supercooled (that is, metastable) liquids such as Y_2O_3–Al_2O_3 mixtures[8], water[9] and other molecular liquids[10–12]. However, the LLT in supercooled liquids often occurs simultaneously with crystallization, making it difficult to separate the two phenomena[13]. A liquid–liquid critical point (LLCP), similar to the gas–liquid critical point, has been predicted at the end of the LLT line that separates the low- and high-density liquids in some cases, but has not yet been experimentally observed for any materials. This putative LLCP has been invoked to explain the thermodynamic anomalies of water[1]. Here we report combined in situ density, X-ray diffraction and Raman scattering measurements that provide direct evidence for a first-order LLT and an LLCP in sulfur. The transformation manifests itself as a sharp density jump between the low- and high-density liquids and by distinct features in the pair distribution function. We observe a non-monotonic variation of the density jump with increasing temperature: it first increases and then decreases when moving away from the critical point. This behaviour is linked to the competing effects of density and entropy in driving the transition. The existence of a first-order LLT and a critical point in sulfur could provide insight into the anomalous behaviour of important liquids such as water.

The pressure–temperature (*P*–*T*) phase diagram of sulfur exhibits important similarities to that of phosphorus, which is so far the only element for which a direct in situ realization of an LLT has been unambiguously demonstrated[14–16]. The stable sulfur solids at ambient pressure (100 kPa), α- and β-sulfur[17,18], consist of S_8 molecules, whereas at high pressure and temperature, the stable polymorph is a polymeric solid composed of helical chains[19]. At room pressure the molecular character is conserved in the liquid upon melting at 388 K and up to 432 K, where the so-called 'λ-transition' occurs[20]. The λ-transition has been described as a 'living' polymerization transition[21], reversible and incomplete (the polymer content reaches a maximum of ~60% at the boiling point, $T = 718$ K), in which a fraction of the S_8 cyclic molecules open up and coalesce into long polymeric chains or rings. It is associated with a large increase in viscosity and an anomalous, but not discontinuous, density variation[20,22]. For $P > 5$ GPa and $T > 1{,}000$ K, several *P*–*T* domains with different thermal and electrical properties have been proposed in liquid sulfur[23]. An experimental study[24] reported at pressures above 6 GPa the existence of a purely polymeric liquid composed of long chains below 1,000 K, which split to shorter chains at higher temperatures. Ab initio molecular dynamics simulations[25] reproduced this chain breakage in the compressed liquid but found no discontinuous change of density associated with this process. So far, no in situ structural or vibrational studies have been conducted in the pressure region below 3 GPa in the mixed molecular–polymeric liquid.

We performed in situ X-ray absorption, X-ray diffraction and Raman scattering measurements at the beamline ID27 of the European Synchrotron Radiation Facility (ESRF)[26] to probe the density, structure and dynamics evolution of liquid sulfur in the *P*–*T* domain at 0–3 GPa and 300–1,100 K (see Supplementary Information sections S1–S3 for the methods). The *P*–*T* paths are presented in the experimental phase diagram of sulfur in Fig. 1.

Density measurements were obtained using a Paris–Edinburgh press along eight isothermal (P1–P8 in Fig. 1) and two isobaric (P9, P10) pathways. The accuracy of the density measured by this method is of the order of 1% (Supplementary Information section S1). X-ray diffraction patterns of the sample were also collected at each *P*–*T* point to confirm that the sample was fully molten. As shown in Fig. 2a, below 1,000 K, along isothermal pathways P1–P5, we systematically observed a discontinuous jump in density over a very narrow pressure range of ~0.07 GPa, which strongly suggests the existence of a first-order phase transition between a low- (LDL) and a high-density liquid (HDL). Discontinuous density shifts were also observed upon varying the

---

[1]European Synchrotron Radiation Facility (ESRF), Grenoble, France. [2]CEA, DAM, DIF, Arpajon, France. [3]Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (IMPMC), Sorbonne Université, CNRS UMR 7590, MNHN, Paris, France. ✉e-mail: mezouar@esrf.fr

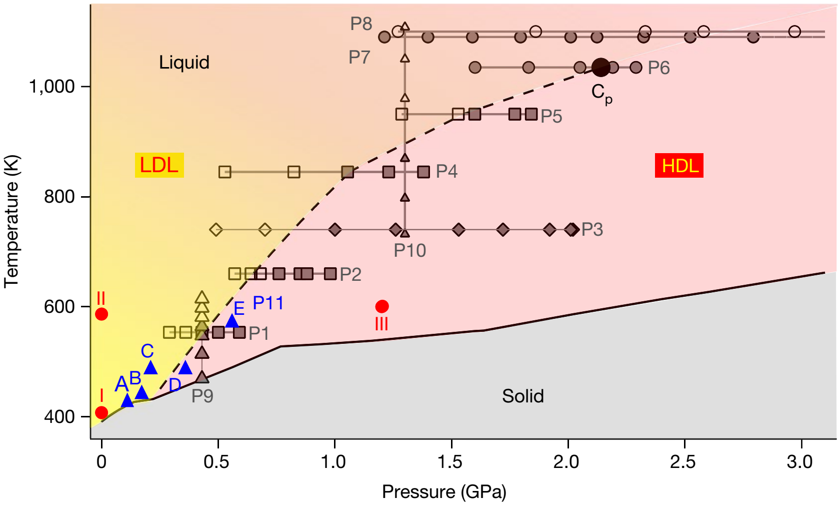


Fig. 1 | Phase diagram of sulfur around the LLT. P1–P8: isothermal pathways followed during the density measurements presented in Fig. 2. P1, P2, P4–P7 were made on compression, whereas P3 (diamonds) and P7 (open black circles) were made on decompression. For clarity, P7 and P8 are shown up to 3 GPa only. P9 and P10 are isobaric pathways followed during the density measurements presented in Supplementary Fig. 9 (Supplementary Information section S1). A, B, C, D and E (blue filled triangles) along path P11 indicate the *P*, *T* conditions of the selected X-ray diffraction data in Fig. 3. I, II and III are the (*P*, *T*) points of the Raman spectra presented in Fig. 3. The black dashed line is the transition line between the LDL domain (yellow) and the HDL domain (pink) that terminates at the critical point C_p (black solid circle).

temperature at constant load along pathways P9, P10 (Supplementary Fig. 9). These density jumps were accompanied by sudden changes in the structure factor $S(Q)$ of the sulfur melt, as shown in Fig. 2b. This is particularly apparent in the width and position of the first diffraction peak, which change abruptly at the transition. The density variation estimated from the measured $S(Q)$ using the methodology of ref. [27] (see Supplementary Information section S3) compares very well with that derived from the X-ray absorption measurements (see inset of Fig. 2c), giving an independent confirmation of the discontinuous density jump at the transition.

Figure 2c presents X-ray radiographic images taken along a compression pathway at 980 K and at pressures of 1.6–2.5 GPa. These images show that below (i) and above (iv) the transition, the sample is homogeneous, whereas at the transition an interface separating a ‘bubble’ of HDL from the surrounding LDL appears (ii and iii). As seen in Supplementary Video 1, the HDL bubble grows as the load is increased, until the sample is fully in the HDL phase. These observations provide compelling evidence of the coexistence of the LDL and HDL phases at the transition and, together with the density and structure factor measurements, confirm the first-order nature of the LLT.

At 1,090 K and 1,100 K (pathways P7 and P8 in Fig. 1), we did not observe (within uncertainties) any discontinuous shift of the density as a function of pressure; this indicates the presence of an LLCP. The LLCP was probably crossed along pathway P6 at ~2.15 GPa and 1,035 K (star in Fig. 2a), where the density measurements show clear anomalies (see Supplementary Information section S2), whereas at lower and higher pressures along this isotherm, the density appears to vary continuously with pressure. As shown in Fig. 2d, we observed a non-monotonic evolution of the density discontinuity with temperature: starting from zero at the LLCP, it first increases to a maximum of ~7.5% at about 750 K, and then decreases.

Figure 3a shows the pair distribution function (PDF) $g(r)$ obtained by Fourier transform of the measured $S(Q)$ at five selected points along the P8 pathway: A, B and C are in the LDL domain, whereas D and E are in the HDL domain, close to the LDL–HDL transition line (see Fig. 1). The LDL PDF at point A (0.11 GPa, 428 K) is very similar to those reported for the ambient-pressure molecular liquid below the λ-transition[28,29]. It has three well defined peaks at 2.05(2), 3.39(2) and 4.45(2) Å (uncertainties indicate 1 s.d.), in very good agreement with the previous cited works. As shown in ref. [29], the third peak is a fingerprint of the S_8 molecule because it occurs at the average distance of the third and fourth neighbours in an S_8 ring, as deduced from the structure of the molecular α-sulfur crystal. When the temperature is increased in the LDL domain above the λ-transition (points B at 0.17 GPa, 442 K and C at 0.36 GPa, 487 K), the observed evolution is also very similar to that described in the literature for the ambient-pressure liquid[28]. Namely, the positions and intensities of the first- and second-neighbour peaks are weakly affected, whereas the third peak is strongly reduced in intensity and becomes bimodal. These changes in the third- and fourth-neighbour distribution are a signature of the rapid increase of polymer content and the associated reduction of the S_8 content above the λ-transition. Indeed, the peak at 4.45 Å, characteristic of S_8 molecules, is reduced in intensity, and a new component, originating from the formation of long polymeric chains or rings, appears at 4 Å and grows with temperature. The similarities between the present PDF in the region from 4 Å to 5 Å and those reported in ref. [28] from ambient-pressure neutron diffraction can be appreciated from the comparison between the inset of Fig. 3a and Fig. 3b. We note, however, that the new component in the ambient-pressure PDF appears at a larger distance, around 4.2 Å, which is probably due to the lower density of the ambient-pressure liquid.

We now come to the structural modifications in the PDF across the LDL–HDL transition. As seen in Fig. 3c and Supplementary Information section S15, no change (within uncertainties) occurs on the first and second peak positions, showing that the S–S bond length and the ⟨S–S–S⟩ angle are the same as in the LDL. The most important modifications occur again in the third- and fourth-neighbour distributions. The bimodal shape of the third peak is maintained but the component at 4.45 Å is even more reduced, and the component located at 4 Å in the LDL undergoes a sudden shift in position to 4.15 Å in the HDL. This shows that the local order in the liquid changes at the transition, and further suggests that the polymer content in the HDL is larger than in the LDL. The latter point is confirmed by the comparison of the Raman spectra measured in the LDL and HDL shown in Fig. 3d. In the HDL domain, we observe an increase in intensity of the stretching mode at 460 cm^{-1} that is assigned to the polymeric chains[30], concomitant with a decrease of the molecular bending mode at 152 cm^{-1} and the breathing mode at 220 cm^{-1} of the S_8 molecule, attesting that the latter are residual in the HDL region.

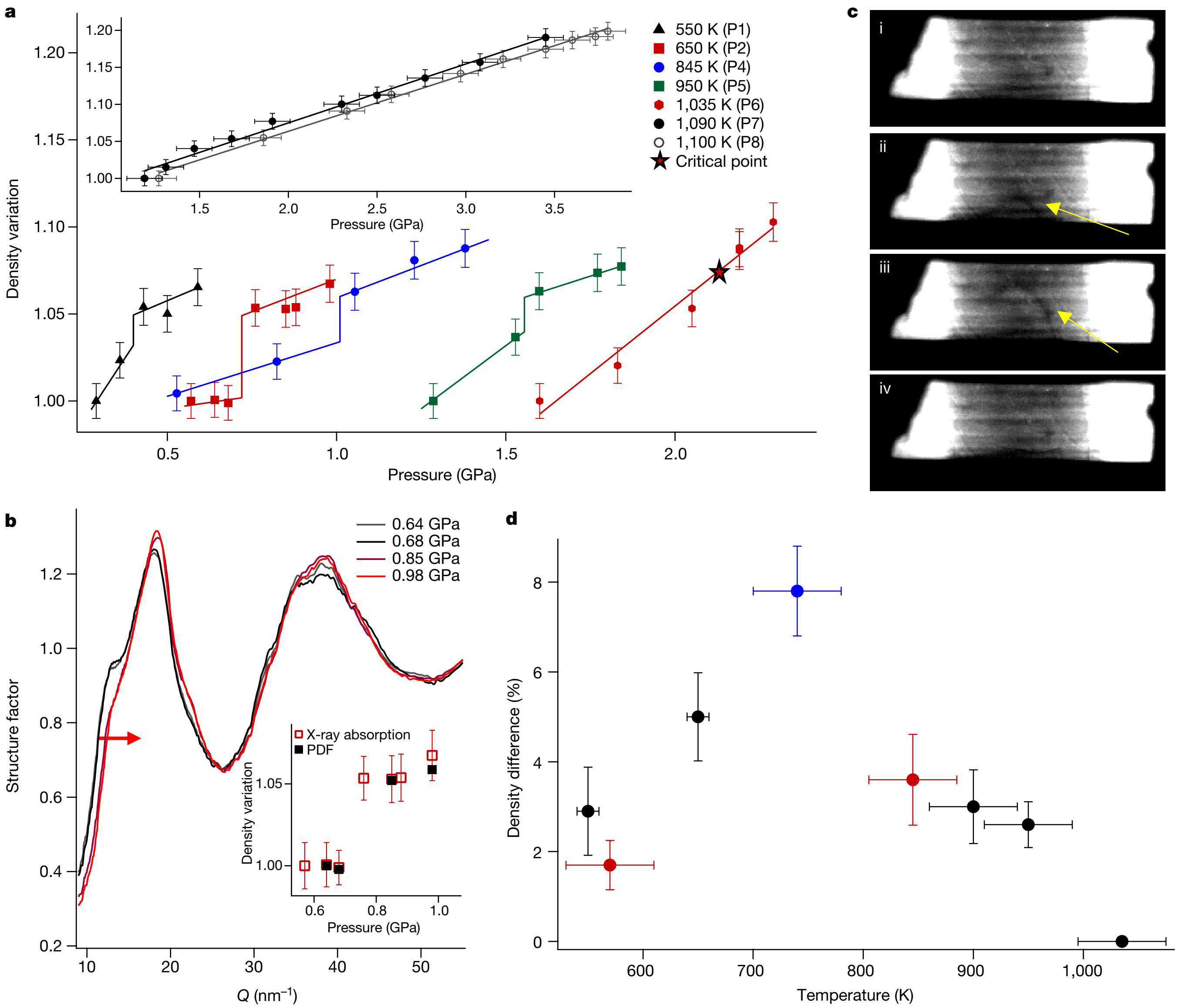


Fig. 2 | First-order LLT in sulfur. a, Relative pressure variation of the liquid density, ρ/ρ_0 (ρ_0 is the density of the lowest pressure point for each isotherm) collected along seven isothermal pathways (P1, P2 and P4–P8 in Fig. 1). For clarity, the density jump obtained on decompression along P3 and along the isothermal paths P9 and P10 are presented in Extended Data Figs. 1, 3. Main panel: at temperatures below 1,030 K, a clear density jump is observed along all the isothermal paths. At ~1,035 K, a density anomaly is detected in the vicinity of the LLCP (see Extended Data Fig. 4). Above the LLCP (inset), a continuous variation of the density is observed. **b**, Structure factors $S(Q)$ of liquid sulfur collected along the isothermal path P2 (T = 650 K). The red arrow emphasizes the shift of the first peak position of $S(Q)$ at the LDL–HDL transition. The $S(Q)$ data collected on decompression at 740 K are shown in Extended Data Fig. 2. The variation of the density calculated from the associated PDF (black filled squares) is presented in the inset together with the one obtained from the direct density measurements (red empty squares). **c**, X-ray radiography of liquid sulfur across the LDL–HDL transition line at T = 980 K and at pressures between 1.6 GPa and 2.5 GPa. Results are shown for pure LDL (i), LDL and HDL coexistence (ii, iii), pure HDL (iv). The yellow arrows indicate the LDL–HDL boundary. **d**, Temperature evolution of the density jump. The black and blue symbols correspond to the isothermal pathways (P1–P8) and the red symbols to the isobaric (P9, P10) pathways. The maximum of 7.5% (±1 s.d.) is located at ~750 K. Error bars indicate 1 s.d.

This work thus demonstrates that sulfur undergoes a first-order phase transition between two thermodynamically stable liquids, with clear experimental evidence of a sharp density increase and structural modifications. We stress that this LLT is distinct from the long-known λ-transition, which is associated with second-order-like changes in density[20,22] and heat capacity[31]. Furthermore, the λ-transition temperature slowly decreases with pressure[32], whereas for the present LLT the transition temperature increases with pressure (see also Supplementary Information section S4). By virtue of the Clapeyron equation, and because a positive jump of density occurs at the transition, this indicates that the entropy of the HDL is smaller than that of the LDL, which contrasts with the entropy increase in the LDL at the λ-transition[33]. The entropy reduction across the LLT may be due in part to the increase in polymer content revealed by these experiments and the associated reduction in the mixing entropy. However, the observed changes in the PDF also indicate that the local conformation of neighbouring polymeric units is modified to a more compact arrangement imposed by the density increase, leading to a reduction in the conformation entropy as well.

Because of the shape of the transition line and the presence of a critical point, this LLT in sulfur strongly resembles the well known liquid–gas transition. However, there is an important difference: the non-monotonic variation of the density jump with temperature of the LLT, which first increases from zero as the temperature is decreased

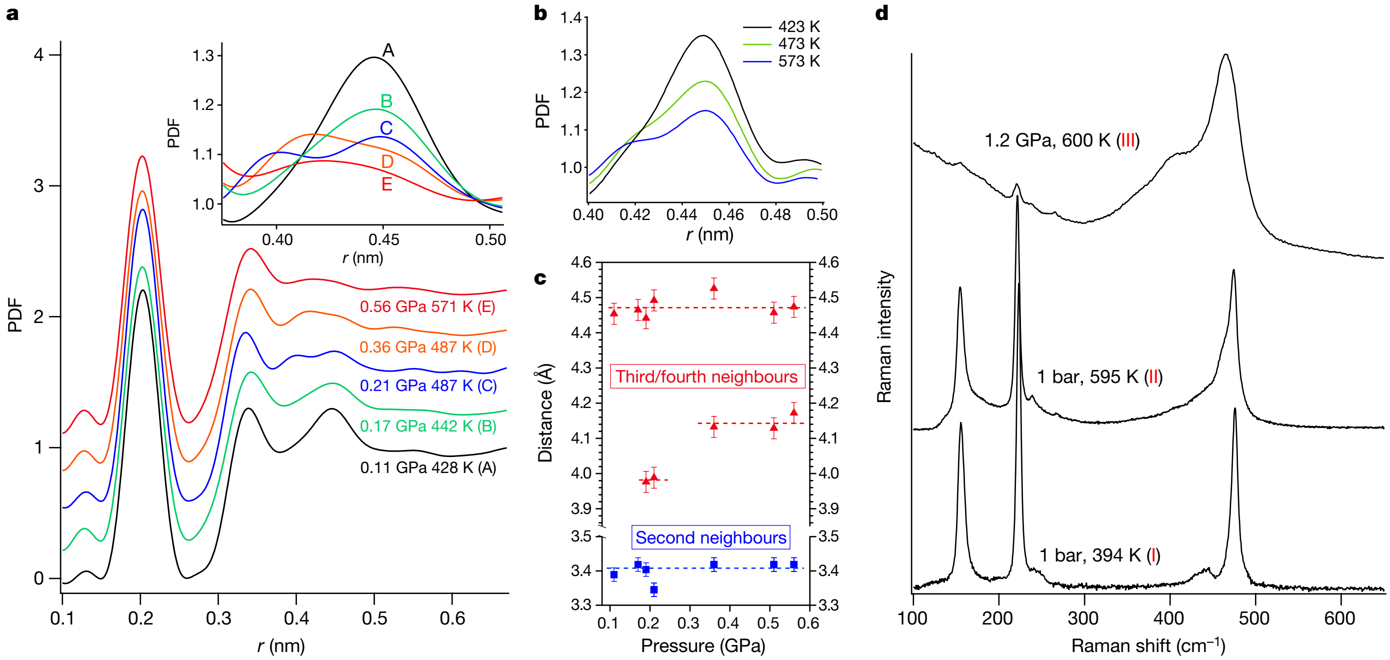


Fig. 3 | Local order in the LDL and HDL sulfur. a, The PDF, *g*(*r*), of liquid sulfur at selected pressure and temperature conditions along path P11 of Fig. 1. Curves A–E correspond to the *P*, *T* conditions in Fig. 1 at which the measurements were performed. The curves are vertically offset by 0.3 for clarity. The inset shows a magnification of the PDF in the region 0.4–0.5 nm, which contains contributions from third and fourth neighbours. **b**, Magnification of the third peak of *g*(*r*) from the ambient-pressure neutron data of ref. [28] at 423 K, 473 K and 573 K. The curves were reproduced with permission from ref. [28] (American Physical Society, 1990). **c**, Positions of the second, third and fourth neighbours as a function of pressure, as deduced from the peak positions of the PDF along path P11. **d**, Selected Raman spectra of liquid sulfur collected at the (*P*, *T*) points I, II and III of Fig. 1: I, LDL below the λ-transition; II, LDL above the λ-transition; III, HDL. Error bars indicate 1 s.d.

from the LLCP, and then decreases, in contrast to the monotonic increase of density along the liquid–gas transition. Such a behaviour was recently predicted by ab initio molecular dynamics calculations along the LLT line of phosphorus[34], which suggested that the order parameter describing the LLT contains contributions from both the density and the entropy and that at least at low temperatures, entropy—rather than density—governs the transition. This is at odds with the liquid–gas transition, for which density is the sole order parameter. A two-order-parameter model including density and a bond-order parameter describing locally favoured structures has been proposed[13] to explain the existence of LLTs. The putative LLT in water has also been described as entropy-driven, on the basis of a model in which water is considered as an 'athermal solution' of two molecular structures with different entropies and densities[35]. We note, however, that the present LLT in sulfur is different from that in water and phosphorus, in the sense that the transition line has a positive slope in sulfur but a negative one in water and phosphorus. This may signal that sulfur belongs to a different class of LLT.

This work also provides the first, to the best of our knowledge, experimental evidence for a critical point terminating the line of an LLT. Such an LLCP was proposed in phosphorus at about 3,500 K and 0.02 GPa (ref. [34]), conditions that have not yet been achieved experimentally. In supercooled liquid silicon, classical empirical calculations have predicted an LLCP at negative pressures (−0.6 GPa, 1,120 K)[36] but ab initio calculations recently determined that the HDL– and HDL–vapour spinodals form a continuous reentrant curve, making supecooled Si a critical-point-free system[37]. In water, the existence of an LLCP was proposed[1] in 1992 to explain the many anomalies in the thermodynamic properties of water, such as the heat capacity, compressibility and thermal expansion coefficients, and evidence for this LLCP has been explored and debated ever since. The experimental observation of this hypothetical LLCP in water may never be possible as it is located in the 'no man's land', that is, the *P*–*T* domain below the homogeneous-nucleation temperature. Finally, an LLCP has also been predicted in simple molecular systems, such as H_2 (ref. [6]) and N_2 (ref. [7]), but their experimental observation remains extremely challenging. The LLCP in sulfur, being in a *P*–*T* range easily accessible by experiment, provides a unique opportunity for the study of critical phenomena associated with LLTs. We thus expect that the present work will generate a new interest in LLTs that will provide a solid basis for understanding the principles that govern LLTs in general. Future studies should also focus on deciphering the microscopic processes at the origin of the LLT in sulfur.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2593-1.

1. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
2. Harrington, S., Zhang, R., Poole, P. H., Sciortino, F. & Stanley, H. E. Liquid–liquid phase transition: evidence from simulations. *Phys. Rev. Lett.* **78**, 2409–2412 (1997).
3. Sastry, S. & Angell, C. A. Liquid–liquid phase transition in supercooled silicon. *Nat. Mater.* **2**, 739–743 (2003).
4. Boates, B., Teweldeberhan, A. M. & Bonev, S. A. Stability of dense liquid carbon dioxide. *Proc. Natl Acad. Sci. USA* **109**, 14808–14812 (2012).
5. Glosli, J. N. & Ree, F. H. Liquid–liquid phase transformation in carbon. *Phys. Rev. Lett.* **82**, 4659–4662 (1999).
6. Morales, M. A., Pierleoni, C., Schwegler, E. & Ceperley, D. M. Evidence for a first-order liquid–liquid transition in high-pressure hydrogen from ab initio simulations. *Proc. Natl Acad. Sci. USA* **107**, 12799–12803 (2010).
7. Boates, B. & Bonev, S. First-order liquid–liquid phase transition in compressed nitrogen. *Phys. Rev. Lett.* **102**, 015701 (2009).
8. Aasland, S. & McMillan, P. F. Density-driven liquid–liquid phase separation in the system Al_2O_3–Y_2O_3. *Nature* **369**, 633–636 (1994).

9. Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
10. Woutersen, S., Ensing, B., Hilbers, M., Zhao, Z. & Angell, C. A. A liquid–liquid transition in supercooled aqueous solution related to the HDA-LDA transition. *Science* **359**, 1127–1131 (2018).
11. Tanaka, H., Hurita, R. & Mataki, H. PRL 92, Liquid–liquid transition in the molecular liquid triphenyl phosphite. *Phys. Rev. Lett*. **92**, 025701–025704 (2004).
12. Kurita, R. & Tanaka, H. On the abundance and general nature of the liquid–liquid phase transition in molecular systems. *J. Phys. Condens. Matter* **17**, 293–302 (2005).
13. Murata, K. & Tanaka, H. Microscopic identification of the order parameter governing liquid–liquid transition in a molecular liquid. *Proc. Natl Acad. Sci. USA* **112**, 5956–5961 (2015).
14. Katayama, Y. et al. First-order liquid–liquid phase transition in phosphorus. *Nature* **403**, 170–173 (2000).
15. Monaco, G., Falconi, S., Crichton, W. A. & Mezouar, M. Nature of the first-order phase transition in fluid phosphorus at high temperature and pressure. *Phys. Rev. Lett*. **90**, 255701 (2003).
16. Katayama, Y. et al. Macroscopic separation of dense fluid phase and liquid phase of phosphorus. *Science* **306**, 848–851 (2004).
17. Steudel, R. & Eckert, B. Solid sulfur allotropes. *Top. Curr. Chem*. **230**, 1–80 (2003).
18. Templeton, L. K., Templeton, D. H. & Zalkin, A. Crystal structure of monoclinic sulfur. *Inorg. Chem*. **15**, 1999–2001 (1976).
19. Crichton, W. A., Vaughan, G. B. M. & Mezouar, M. In situ structure solution of helical sulfur at 3 GPa and 400C. *Z. Kristallogr*. **216**, 417–419 (2001).
20. Sauer, G. E. & Borst, L. B. Lambda transition in liquid sulfur. *Science* **158**, 1567–1569 (1967).
21. Tobolsky, A. V. & Eisenberg, A. Equilibrium polymerization of sulfur. *J. Am. Chem. Soc*. **81**, 780–782 (1959).
22. Zheng, K. M. & Greer, S. C. The density of liquid sulfur near the polymerization temperature. *J. Chem. Phys*. **96**, 2175–2182 (1992).
23. Brazhkin, V. V., Popova, S. V. & Voloshin, R. N. Pressure–temperature phase diagram of molten elements: selenium, sulfur and iodine. *Physica B* **265**, 64–71 (1999).
24. Liu, L. et al. Chain breakage in liquid sulfur at high pressures and high temperatures. *Phys. Rev. B* **89**, 174201 (2014).
25. Plašienka, D., Cifra, P. & Martoňák, R. Structural transformation between long and short-chain form of liquid sulfur from ab initio molecular dynamics. *J. Chem. Phys*. **142**, 154502–154512 (2015).
26. Mezouar, M. et al. Development of a new state-of-the-art beamline optimized for monochromatic single-crystal and powder X-ray diffraction under extreme conditions at the ESRF. *J. Synchrotron Rad*. **12**, 659–664 (2005).
27. Eggert, J., Weck, G., Loubeyre, P. & Mezouar, M. Quantitative structure factor and density measurements of high-pressure fluids in diamond anvil cells by X-ray diffraction: argon and water. *Phys. Rev. B* **65**, 174105 (2002).
28. Bellissent, R., Descotes, L., Boué, F. & Pfeuty, P. Liquid sulfur: local-order evidence of a polymerization transition. *Phys. Rev. B* **41**, 2135–2138 (1990).
29. Vahvaselkä, K. S. & Mangs, J. M. X-Ray diffraction study of liquid sulfur. *Phys. Scr*. **38**, 737–741 (1988).
30. Kalampounias, A. G., Kastrissios, D. T. & Yannopoulos, S. N. Structure and vibrational modes of sulfur around the λ-transition and the glass transition. *J. Non-Cryst. Solids* **326–327**, 115–119 (2003).
31. Braune, H. & Moller, O. The specific heat of liquid sulfur. *Z. Naturforsch. B* **9a**, 210–217 (1954).
32. Kuballa, M. & Schneider, G. Differential thermal analysis under high pressure I: investigation of the polymerisation of liquid sulfur. *Ber. Bunsenges. Phys. Chem* **75**, 513–516 (1971).
33. Steudel, R. Liquid sulfur. *Top. Curr. Chem*. **230**, 81–116 (2003).
34. Zhao, G. et al. Anomalous phase behavior of first-order fluid–liquid phase transition in phosphorus. *J. Chem. Phys*. **147**, 204501 (2017).
35. Holten, V. & Anisimov, M. A. Entropy-driven liquid–liquid separation in supercooled water. *Sci. Rep*. **2**, 713 (2012).
36. Vasisht, V. V., Saw, S. & Sastry, S. Liquid–liquid critical point in supercooled silicon. *Nat. Phys*. **7**, 549–553 (2011).
37. Zhao, G. et al. Phase behavior of metastable liquid silicon at negative pressure: ab initio molecular dynamics. *Phys. Rev. B* **93**, 140203 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Data availability

The data that support the findings of this study are available from the corresponding author upon request. Source data are provided with this paper.

Acknowledgements We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron beamtime at beamline ID27, the Agence Nationale de la Recherche for financial support under grant number ANR 13-BS04-0015 (MOFLEX) and Almax easy Lab for providing the diamond cylinders.

Author contributions The original idea was conceived by M.M. The experiments were performed by L.H., G.G., D.S. and M.M. with equal contributions. The data were analysed and the figures produced by L.H. with contributions from all the co-authors. The manuscript was written by M.M. and F.D. with contributions from all the co-authors.

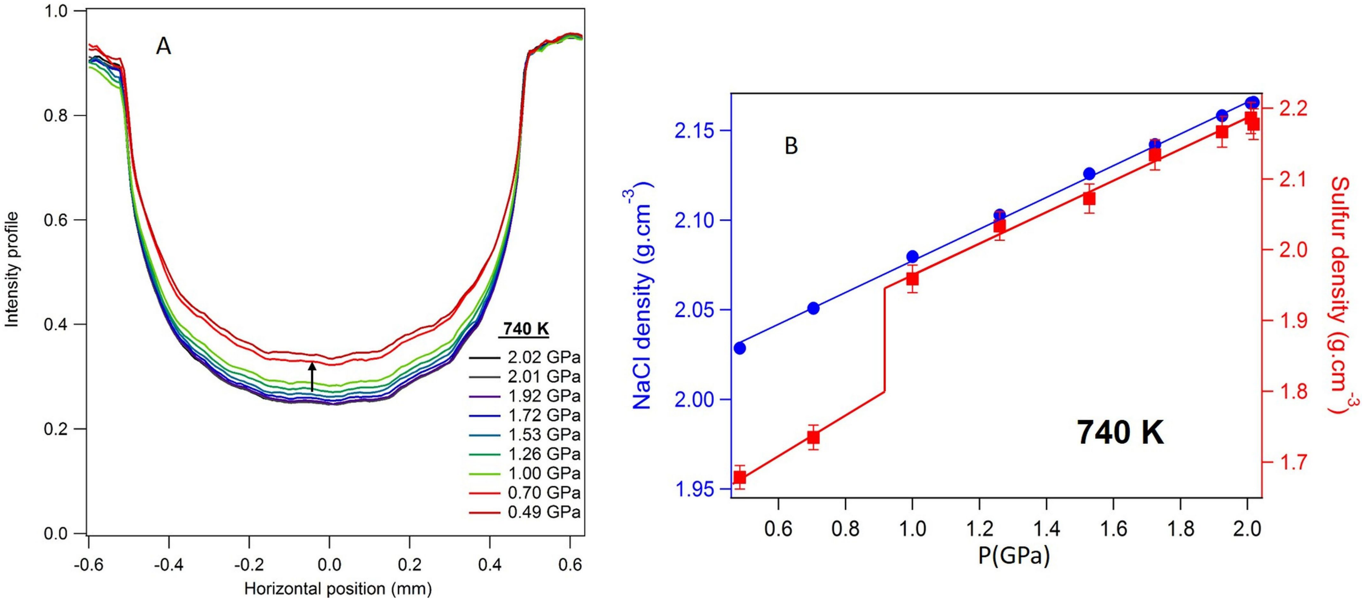
Competing interests The authors declare no competing interests.

Additional information

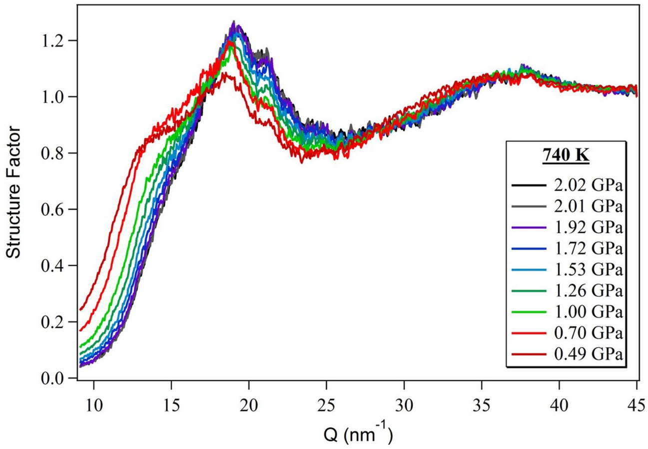
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2593-1.

Correspondence and requests for materials should be addressed to M.M.

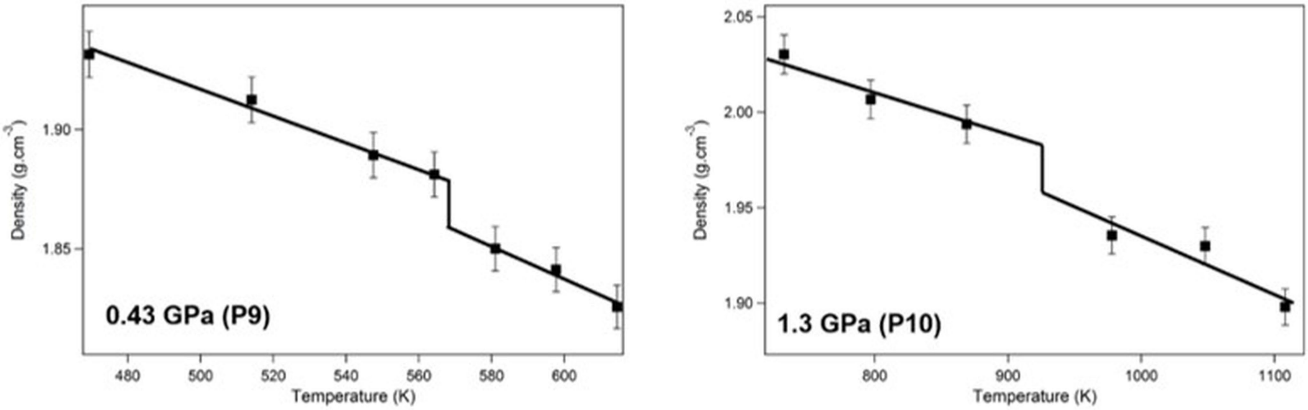
Peer review information *Nature* thanks Yoshio Kono, Wenge Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



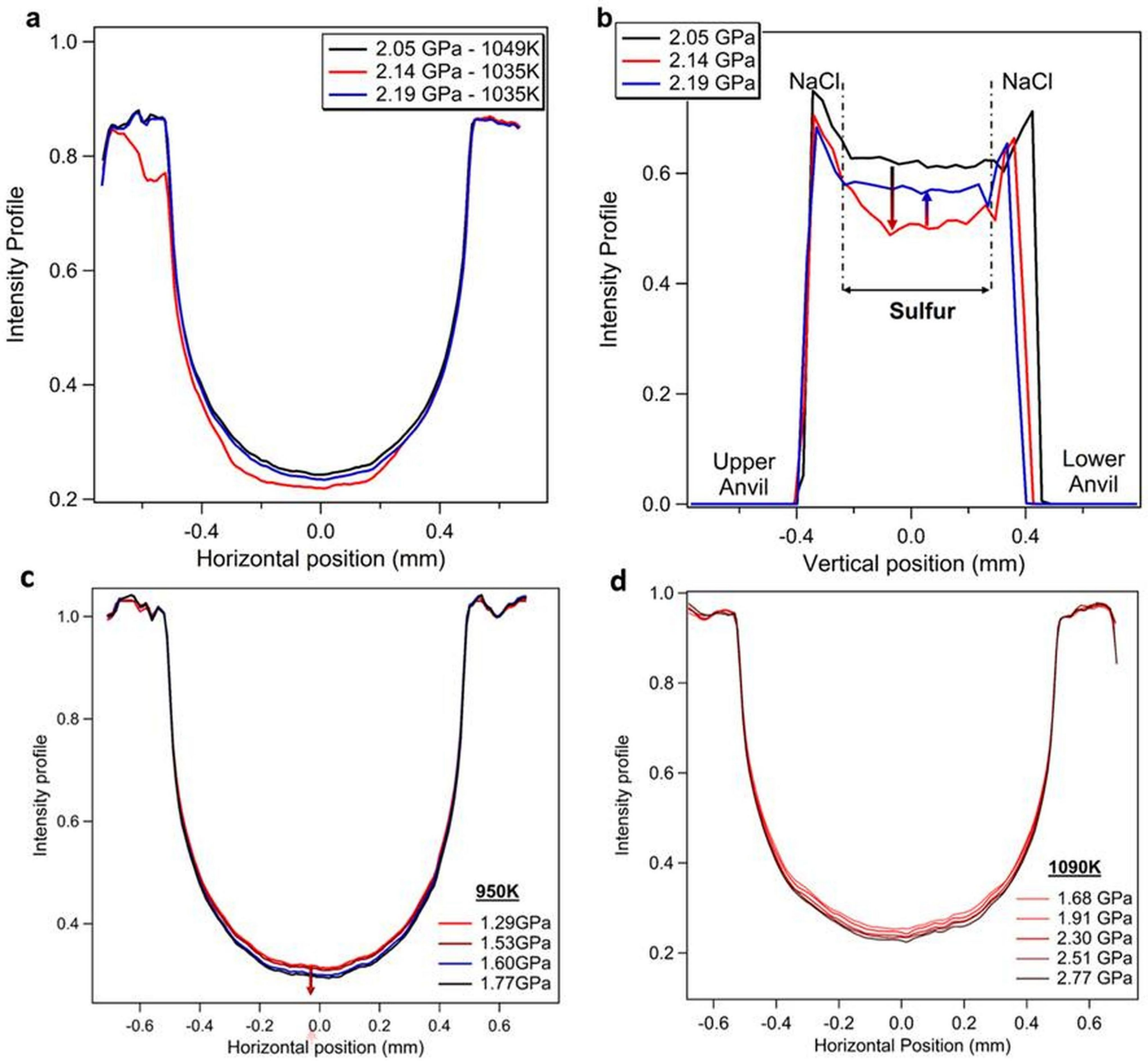
Extended Data Fig. 1 | Density discontinuity at 740 K. **a**, Raw datasets of isothermal X-ray absorption profiles I/I_0 (where I_0 and I are the incident and transmitted intensities, respectively) collected on decompression at 740 K. The black arrow indicates the density jump. **b**, Resulting isothermal density curve of sulfur (red) and density variation of NaCl pressure standard (blue). Error bars indicate 1 s.d.



Extended Data Fig. 2 | Structure factors. Structure factors (S(Q)) of liquid sulfur collected on decompression along the isothermal path at T = 740 K.



Extended Data Fig. 3 | Isothermal density discontinuity. Density of liquid sulfur as a function of temperature along isobaric paths P9 at 0.4 GPa (left) and P10 at 1.3 GPa (right). Error bars indicate 1 s.d.



Extended Data Fig. 4 | LLCP in sulfur. a, **b**, X-ray absorption profiles I/I_0 in the horizontal (**a**) and vertical (**b**) directions in the vicinity of the critical point. During the measurements, the X-ray beam was stopped by the upper and lower anvils of the Paris–Edinburgh press. **c**, **d**, Horizontal X-ray absorption profiles at temperatures below (**c**; 950 K) and above (**d**; 1,090 K) the critical point. The red arrow in **c** indicates the I/I_0 discontinuity at the LLT. No I/I_0 discontinuity is observed at temperatures above the critical point (**d**).

Additive manufacturing of silica aerogels

https://doi.org/10.1038/s41586-020-2594-0

Received: 11 November 2019

Accepted: 28 May 2020

Shanyu Zhao[1,9]✉, Gilberto Siqueira[2,9], Sarka Drdova[3,4], David Norris[1], Christopher Ubert[1], Anne Bonnin[5], Sandra Galmarini[1], Michal Ganobjak[1,6], Zhengyuan Pan[3,7], Samuel Brunner[1], Gustav Nyström[2,8], Jing Wang[3,4], Matthias M. Koebel[1] & Wim J. Malfait[1]✉

Owing to their ultralow thermal conductivity and open pore structure[1–3], silica aerogels are widely used in thermal insulation[4,5], catalysis[6], physics[7,8], environmental remediation[6,9], optical devices[10] and hypervelocity particle capture[11]. Thermal insulation is by far the largest market for silica aerogels, which are ideal materials when space is limited. One drawback of silica aerogels is their brittleness. Fibre reinforcement and binders can be used to overcome this for large-volume applications in building and industrial insulation[5,12], but their poor machinability, combined with the difficulty of precisely casting small objects, limits the miniaturization potential of silica aerogels. Additive manufacturing provides an alternative route to miniaturization, but was “considered not feasible for silica aerogel”[13]. Here we present a direct ink writing protocol to create miniaturized silica aerogel objects from a slurry of silica aerogel powder in a dilute silica nanoparticle suspension (sol). The inks exhibit shear-thinning behaviour, owing to the high volume fraction of gel particles. As a result, they flow easily through the nozzle during printing, but their viscosity increases rapidly after printing, ensuring that the printed objects retain their shape. After printing, the silica sol is gelled in an ammonia atmosphere to enable subsequent processing into aerogels. The printed aerogel objects are pure silica and retain the high specific surface area (751 square metres per gram) and ultralow thermal conductivity (15.9 milliwatts per metre per kelvin) typical of silica aerogels. Furthermore, we demonstrate the ease with which functional nanoparticles can be incorporated. The printed silica aerogel objects can be used for thermal management, as miniaturized gas pumps and to degrade volatile organic compounds, illustrating the potential of our protocol.

Aerogels are mesoporous sol–gel materials with high specific surface area ($500–1{,}000\ \mathrm{m^2\,g^{-1}}$) and low density ($0.001–0.200\ \mathrm{g\,cm^{-3}}$), and are classified as superinsulators owing to their ultralow thermal conductivity (down to $12\ \mathrm{mW\,m^{-1}\,K^{-1}}$). Silica aerogel is by far the most studied and used type of aerogel. It is available in bulk quantities for industrial and building insulation[5,12], with a rapidly growing market of around US$220 million per year[14]. Although aerogels can have exceptionally high strength-to-weight ratios[2], silica aerogels are generally brittle and impossible to machine by subtractive processing. The viability of additive manufacture of aerogels has been demonstrated for graphene[15,16], graphene oxide[17,18], carbon nitride[19], gold[20], resorcinol-formaldhyde[21] and cellulose[22–24], but is “fraught with experimental difficulties and probably not feasible for silica aerogels”[13]. Silica particles are a common additive for 3D printing[25], but an additive manufacturing protocol for pure silica aerogel has not been established. A recent study on biopolymer–silica hybrid aerogels[26] found that they had poor shape fidelity. In addition, the biopolymer additives were retained in the final product, resulting in limited temperature stability and high thermal conductivities.

We print pure silica aerogel objects by direct ink writing (Fig. 1a–e, Extended Data Fig. 1) of a slurry of silica aerogel powder (IC3100, Cabot; particle size, 4–20 μm; Fig. 1f) in a 1-pentanol-based silica sol. The low vapour pressure of pentanol (18 times lower than that of water at 20 °C) prevents drying-induced surface damage, even when printing for extended durations (Extended Data Table 1). The added cost of using industrial-grade silica aerogel powder is negligible for miniaturized applications. The high loading of gel particles (at least 40 vol%) means the ink exhibits the shear thinning behaviour required for direct ink writing (Fig. 1g, h, Extended Data Fig. 2). The ink consists of aerogel grains roughly 10 μm in diameter suspended in a sol with silica nanoparticles around 10 nm in diameter. The rheological behaviour is complex: strain overshoot at large-amplitude shear, as is typical for colloidal suspensions[27], and non-Newtonian shear thinning at small-to-medium-amplitude oscillatory shear, as is typical for suspensions of large particles. The addition of poly(propylene glycol) bis(2-aminopropyl ether) increases the viscosity of the ink, preventing solid–liquid phase separation, and improves its homogeneity

---

[1]Laboratory for Building Energy Materials and Components, Swiss Federal Laboratories for Materials Science and Technology, Empa, Dübendorf, Switzerland. [2]Cellulose and Wood Materials Laboratory, Swiss Federal Laboratories for Materials Science and Technology, Empa, Dübendorf, Switzerland. [3]Institute of Environmental Engineering, ETH Zurich, Zürich, Switzerland. [4]Laboratory for Advanced Analytical Technologies, Swiss Federal Laboratories for Materials Science and Technology, Empa, Dübendorf, Switzerland. [5]Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland. [6]Faculty of Architecture, Slovak University of Technology in Bratislava, Bratislava, Slovakia. [7]State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou, China. [8]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland. [9]These authors contributed equally: Shanyu Zhao, Gilberto Siqueira. ✉e-mail: shanyu.zhao@empa.ch; wim.malfait@empa.ch

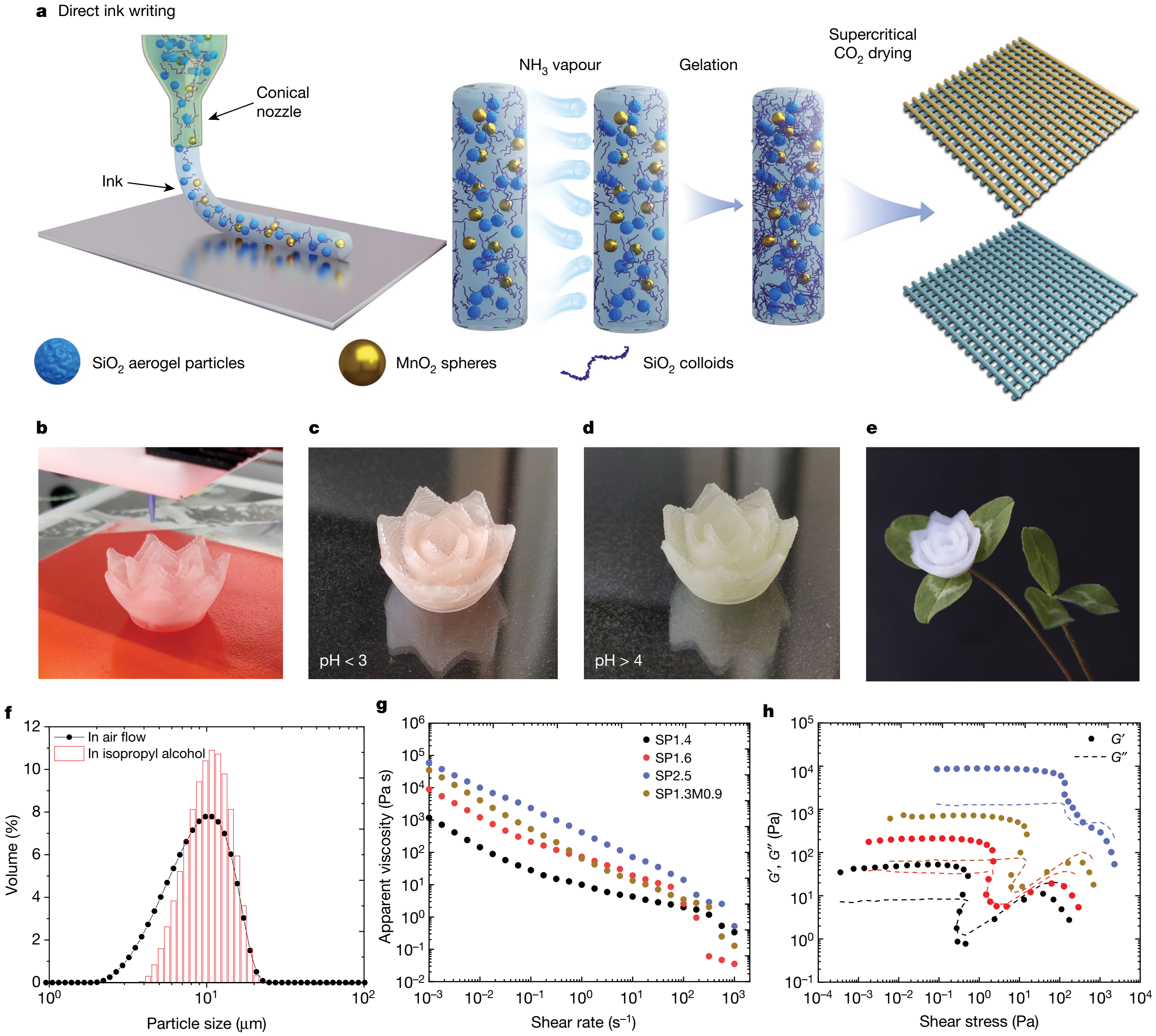


Fig. 1 | Additive manufacture of silica aerogel by direct ink writing. **a**, Scheme for direct ink writing of silica aerogels. The inks, either neat (blue) or functionalized with, for example, MnO_2 nanoparticles (gold), are printed pneumatically through micronozzles. The printed objects are gelled by a vapour-based pH change, and dried from supercritical CO_2. **b**, 3D lotus flower of silica gel printed from ink SP2.5 (Extended Data Table 1) through a conical nozzle with an inner diameter of 410 μm, with a flow rate of 15 mm s^{-1}. The flower is printed using 38 layers; printing took 26 min. Supplementary Video 1 shows a high-speed version of the printing process. **c**–**e**, Photographs of the hydrogel from **b** before solidification (**c**), after ammonia-vapour-induced gelation (**d**) and after drying (**e**). **f**, Particle size distribution for the silica aerogel. **g**, Shear-thinning behaviour of different inks. **h**, Storage (G') and loss (G'') modulus versus shear stress for the different inks.

during the sol–gel transition (Extended Data Fig. 3a). The ink has a shelf life of more than 20 days (Extended Data Fig. 2e, f). During printing, the ink flows easily through the nozzle because of shear thinning, but the filament retains its shape after printing because of the rapid viscosity increase in the absence of shear. Objects have been printed with filament and nozzle diameters as low as 100 μm (Extended Data Fig. 3b). Smaller diameters should be feasible, given the particle size of the aerogel powder (Fig. 1f), provided the printing system can operate at sufficiently high pressure. A silica sol, incorporated in the ink before printing, is activated with ammonia vapour after the object has been printed to bind the aerogel particles and fill the interstitial voids with silica gel. The printed gel may optionally be hydrophobized before the solvent is removed by supercritical CO_2 drying. MnO_2 (ramsdellite) microspheres are mixed into some of the inks to illustrate the ease of functionalization (light absorption or photothermal catalysis)[28,29].

We printed various aerogel objects with high shape fidelity and precision (Fig. 2a–c, e), including honeycombs, 3D lattices and multi-layered continuous membranes. The printed filaments retain a circular cross-section with a well-defined diameter (for example, 327 ± 6 μm). The rheology of the ink can be adapted to the application: higher viscosity for open structures with large overhangs (up to 45°) and wide spans (for example, 10 mm for a filament with a diameter of 400 μm) (Fig. 2d); lower viscosity to enable filaments to merge into continuous membranes without voids (Fig. 2e). The original aerogel grains are embedded in, and infused by, a low-density aerogel matrix derived from

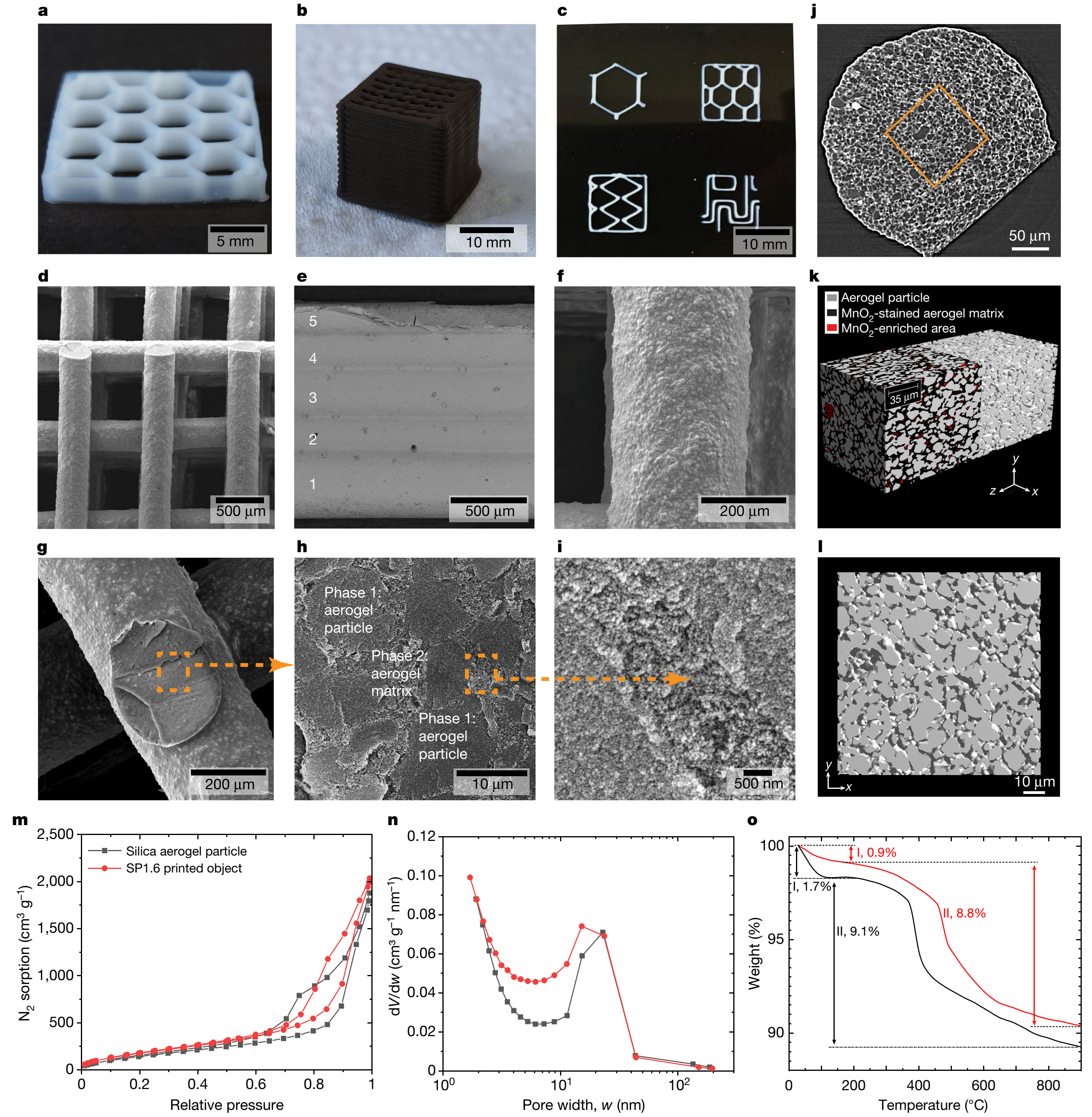


Fig. 2 | 3D-printed objects, their microstructure and selected properties. **a**, A 10-layer honeycomb (ink SP1.6 (Extended Data Table 1), 410-μm conic nozzle, 2.4 min at 15 mm s^{-1}). **b**, A 33-layer lattice cube (ink SP1.3M0.9, 410-μm conic nozzle, 8 min at 12 mm s^{-1}; Supplementary Video 2). **c**, Various 3D patterns (three layers, ink SP1.6, 250-μm conic nozzle, 1 min at 18.4 mm s^{-1}; Supplementary Video 3). **d**, Scanning electron microscopy (SEM) image of a lattice. **e**, A multi-layer continuous membrane. **f**, Outer surface of a printed filament. **g**, Interface between two filaments. **h**, Magnification of the oranged-boxed region in **g**, showing interlocked aerogel particles (darker grey) embedded in a low-density aerogel matrix (lighter grey). **i**, Magnification of the orange-boxed region in **h**, highlighting the two aerogel phases. **j**, Tomographic slice of a printed aerogel filament (ink SP1.6, 410 μm). **k**, 3D volume rendering of the oranged-boxed area in **j** (98 μm × 98 μm × 234 μm), with the low-density aerogel matrix shown (left) and removed (right). Both renderings are stacks of 720 cross-sections. **l**, An x–y section (98 μm × 98 μm) with the low-density matrix removed (darker grey, cut surface of particles; lighter grey, underlaying particle surfaces). **m**, N$_2$ sorption isotherms at 77 K. **n**, Pore size distribution derived from Barrett–Joyner–Halenda (BJH) analysis (V, pore volume; w, pore width). **o**, Thermogravimetric analysis. Weight change I corresponds to adsorbed water, weight change II to trimethylsilyl groups.

the silica sol. These denser grains (Fig. 2f–i) form an interlocked particle packing with direct particle contacts (Fig. 2j–l, Extended Data Fig. 4).

The printed objects consist entirely of hydrophobic silica aerogel, as evidenced by solid-state nuclear magnetic resonance (NMR) and Fourier transform infrared (FTIR) spectroscopy (Extended Data Fig. 5), indicating that the viscosity modifier is fully washed out during the solvent exchanges and/or supercritical drying. The contact angle with water is 150° ± 2°. The high- and low-density aerogel phases display a

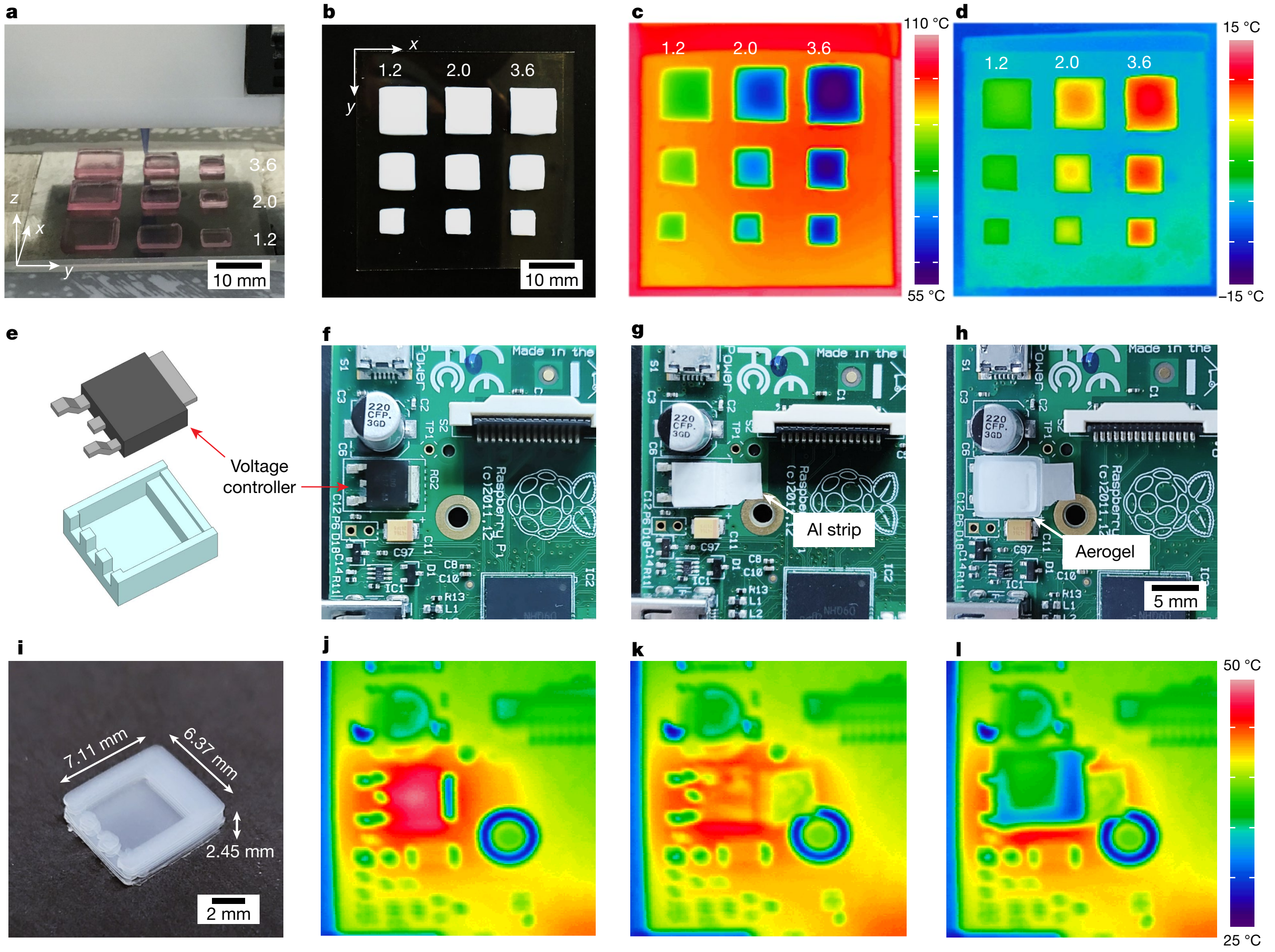


Fig. 3 | Thermal management. a, Design and 3D printing of the gel array (ink SP1.6, 210-µm conic nozzle, 3.5 min at 15 mm s^{-1}; Supplementary Video 4). The thickness of each object is labelled (in millimetres). **b**, Optical image. **c**, Infrared image when placed on a hotplate after more than 0.5 h of equilibration. **d**, Infrared image when placed on an ice block after more than 0.5 h of equilibration. The thickness (in millimetres) of the objects is also labelled in **b**–**d**. **e**, **i**, Sketch (**e**) and photograph (**i**) of a printed aerogel component (410-µm conic nozzle, 2 min at 12 mm s^{-1}). **f**–**h**, **j**–**l**, Photographs (**f**–**h**) and infrared images (**j**–**l**) of a circuit board with neither a sink nor an insulator on the voltage regulator (RG2; **f**, **j**), with an aluminium strip as heat sink (**g**, **k**), and with both a sink and an insulator (**h**, **l**).

network of linked secondary silica particles and high mesoporosity (Fig. 2i). The nitrogen sorption isotherms of the silica aerogel powder starting material and the printed objects are similar; the specific surface area and average pore sizes increase from 697 m^2 g^{-1} to 751 m^2 g^{-1} and from 11.8 nm to 12.6 nm, respectively (Fig. 2m, n). The relatively high bulk density (0.18 ± 0.02 g cm^{-3}) is related to the growth of a low-density silica aerogel phase throughout the entire object—not only in between, but also inside the mesopores of the aerogel powder starting material, owing to the infiltration of the 1-pentanol sol (Extended Data Fig. 1). The high mesopore volume (3.13 cm^3 g^{-1}) limits gas-phase conduction (Knudsen effect). The printed aerogel has a thermal conductivity of 15.9 mW m^{-1} K^{-1} at 25 °C, well below that of standing air (26 mW m^{-1} K^{-1}) or conventional insulation materials (>30 mW m^{-1} K^{-1}). This value is typical for silica aerogel thermal superinsulators, but far lower than that for any 3D-printed object reported so far. Silica aerogel has no lower temperature limit for cryogenic applications. At higher temperatures, radiative conduction increases; the thermal conductivity is expected[30] to increase to around 30 mW m^{-1} K^{-1} at 200 °C and 70 mW m^{-1} K^{-1} at 500 °C. Surprisingly, the printed aerogel has a higher thermal stability than the silica aerogel powder starting material (Fig. 2o). Substantial loss of hydrophobic groups occurs only above 400 °C. The printed samples have similar compressive and tensile strengths as standard silica aerogels, but a better machinability (Extended Data Fig. 6), possibly because the structure with high-density aerogel particles inside a lower-density matrix limits crack propagation. If higher mechanical strength is required, polymer reinforcement has been shown[31] to increase the Young's modulus by a factor of nine and the maximum compressive strength by a factor of seven (Extended Data Fig. 6).

The ability to precisely and reproducibly print superinsulating silica aerogel objects in different sizes and geometries enables new insulation applications. As a demonstration, we printed aerogel features of variable size and thickness onto a substrate (Fig. 3a, b). Thermal imaging when placed onto a hotplate (150 °C) or iceblock (−20 °C) reveals temperature variations that relate directly to the thickness of the printed aerogel insulator (Fig. 3c, d). Combined with appropriately placed heat conductors and sinks, the ultralow thermal conductivity of silica aerogel and the ease with which complex geometries can be produced provide new opportunities for thermal management. Key opportunities could include situations where space is limited, where local hotspots may influence sensitive components or cause damage, and where local temperature gradients need to be restricted, such as in implants, wearable devices, micro-electromechanical systems,

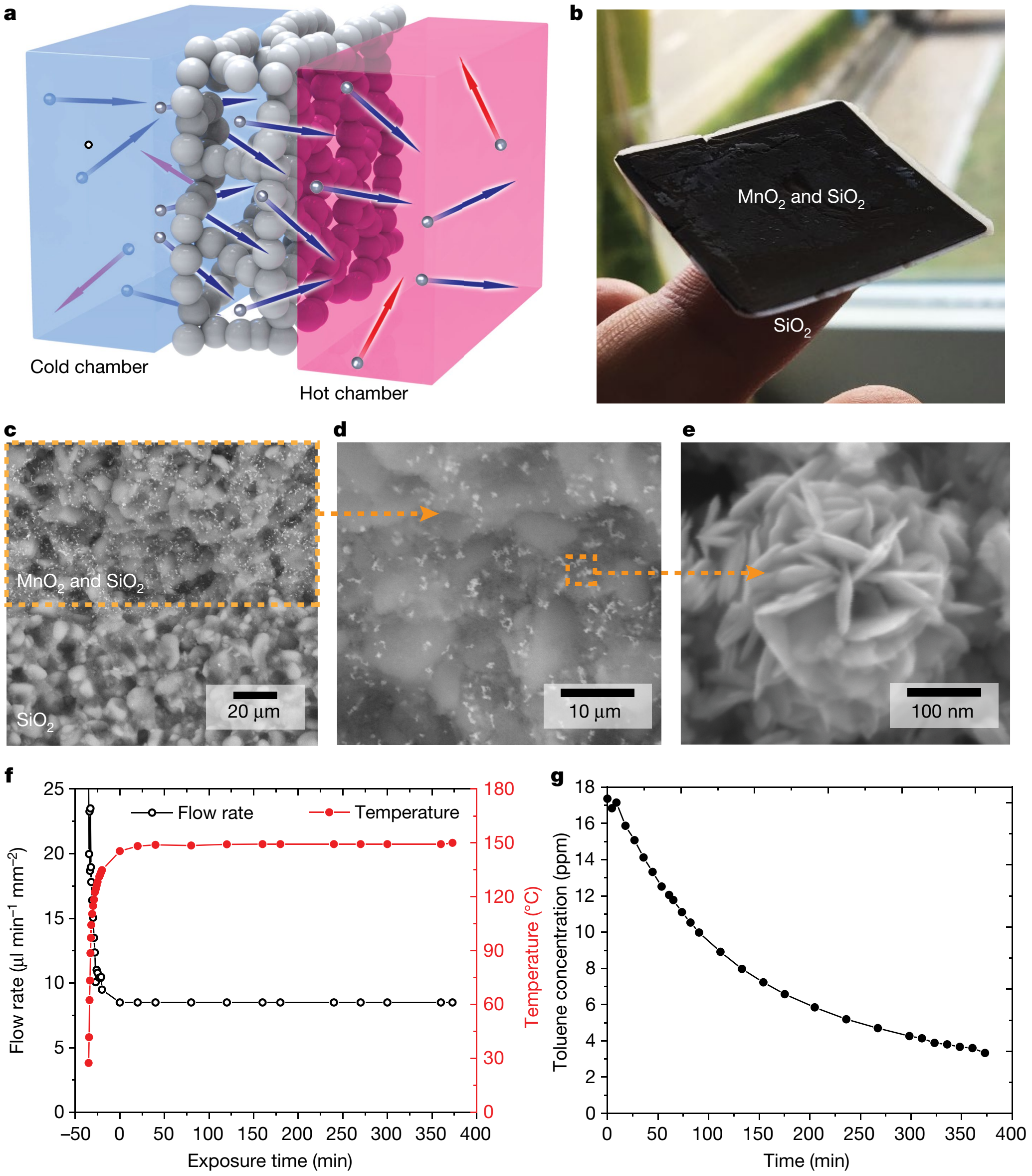


Fig. 4 | Light-driven thermal transpiration gas pump with simultaneous VOC degradation. a, Thermal transpiration. A temperature gradient induces the movement of gas molecules from the cold (blue) to the hot (red) side of a mesoporous membrane. **b**, A bilayer silica aerogel membrane (inks SP1.2 and SP1.3M0.9 (Extended Data Table 1), 410-µm conical nozzle, 6 min at 15 mm s^{-1}). **c**, SEM image of the interface between the two inks. **d**, Magnification of the orange-boxed region in **c**, showing the MnO$_2$ distribution in the silica aerogel. **e**, Magnification of the orange-boxed region in **d**, showing the MnO$_2$ particles. **f**, Light-driven pumping performance. The initial spike in flow rate is due to thermal expansion of gases in the sample chamber; the steady-state flow rate of around 8 µl min^{-1} mm^{-2} is due to thermal transpiration. **g**, Photocatalytic degradation of toluene during thermal transpiration.

smartphones and optical devices. To illustrate the utility of silica aerogel in thermal management, we show that heat can be isolated at the source. A printed aerogel insulator cap, combined with a heat sink, mitigates the local hotspot on a circuit board, making the heat-generating component safe to touch (Fig. 3e–l). We also show that heat-sensitive components can be protected. The local temperature of a capacitor exposed to contact heat is only 36 °C with printed aerogel cap, compared to 75 °C without protection and 48 °C with a cap made from a conventional insulator of the same thickness (Extended Data Fig. 7).

Another application involves using aerogel membranes as a thermal transpiration gas pump. Thermal transpiration generates a gas flow when a thermal gradient is applied to a capillary with a diameter that approaches the mean free path length of the gas molecules (that is, with a Knudsen number K_n that approaches 1; Fig. 4a)[32]. Silica aerogels are ideal membrane materials for Knudsen pumping, owing to their high mesopore volume and low thermal conductivity, which ensures that a steep thermal gradient can be maintained[33]. We printed thin silica aerogel membranes ($K_n = 2.04$; see Methods) with a top layer containing (black) ramsdellite MnO$_2$ microspheres (Fig. 4b–e, Extended Data Fig. 8). Upon light radiation (Extended Data Fig. 9), the black-MnO$_2$-bearing side of the membrane heats up, a thermal-transpiration-driven gas flow is established across the membrane (Fig. 4f), and volatile organic compounds (VOCs; such as toluene) that are part of the gas stream are degraded photothermocatalytically by the MnO$_2$ particles (Fig. 4g).

In summary, our additive manufacturing protocol produces silica aerogel objects with high precision and shape fidelity, the flexibility to include additional functionality and excellent material properties, most notably an ultralow thermal conductivity and high mesoporosity. The 3D-printing process avoids the issues of subtractive manufacturing and opens up new applications for silica aerogels. For thermal insulation, additive manufacturing will enable miniaturized applications (such as portable devices and consumer electronics), adding to the existing silica aerogel market in industrial and building insulation. In addition, the ease with which particulate or polymeric functionality can be incorporated into the ink—illustrated here by the MnO_2-modified inks for thermal transpiration—combined with multi-material printing, enables the production of objects with spatially varying function. This brings electrical, magnetic, optical, chemical and medical applications of silica aerogels into reach, and will allow aerogel phases (along with their tunable functionality) to be integrated into advanced multi-material architectures.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2594-0.

1. Kistler, S. S. Coherent expanded aerogels and jellies. *Nature* **127**, 741 (1931).
2. Hüsing, N. & Schubert, U. Aerogels—airy materials: chemistry, structure, and properties. *Angew. Chem. Int. Ed*. **37**, 22–45 (1998).
3. Prakash, S. S., Brinker, J. C., Hurd, A. J. & Rao, S. M. Silica aerogel films prepared at ambient pressure by using surface derivatization to induce reversible drying shrinkage. *Nature* **374**, 439–443 (1995); erratum **375**, 431 (1995).
4. Wordsworth, R., Kerber, L. & Cockell, C. Enabling Martian habitability with silica aerogel via the solid-state greenhouse effect. *Nat Astron*. **3**, 898–903 (2019).
5. Koebel, M. M., Rigacci, A. & Achard, P. in *Aerogels Handbook* (eds Aegerter, M. et al.) 607–633 (Springer, 2011).
6. Morris, C. A., Anderson, M. L., Stroud, R. M., Merzbacher, C. I. & Rolison, D. R. Silica sol as a nanoglue: flexible synthesis of composite aerogels. *Science* **284**, 622–624 (1999).
7. Kim, C. et al. Large-scale nanoporous metal-coated silica aerogels for high SERS effect improvement. *Sci. Rep*. **8**, 15144 (2018).
8. Pollanen, J. et al. New chiral phases of superfluid ^{3}He stabilized by anisotropic silica aerogel. *Nat. Phys*. **8**, 317–320 (2012).
9. Dumée, L. F. et al. Silver metal nano-matrixes as high efficiency and versatile catalytic reactors for environmental remediation. *Sci. Rep*. **7**, 45112 (2017).
10. Shin, D. et al. Scalable variable-index elasto-optic metamaterials for macroscopic optical components and devices. *Nat. Commun*. **8**, 16090 (2017).
11. Sandford, S. A. et al. Organics captured from Comet 81P/Wild 2 by the Stardust spacecraft. *Science* **314**, 1720–1724 (2006).
12. Lee, K. P., Gould, G. L., Gronemeyer, W. & Stepanian, C. J. Methods to produce gel sheets. US patent 7,399,439 B2 (2008).
13. Bertino, M. F. Rapid fabrication of hybrid aerogels and 3D printed porous materials. *J. Sol-Gel Sci. Technol*. **86**, 239–254 (2018).
14. Collins, R. Aerogels 2019–2029: technologies, markets and players. *IDTechX*, https://www.idtechex.com/en/research-report/aerogels-2019-2029-technologies-markets-and-players/644 (2018).
15. Zhang, Q. et al. 3D printing of graphene aerogels. *Small* **12**, 1702–1708 (2016).
16. Zhu, C. et al. Highly compressible 3D periodic graphene aerogel microlattices. *Nat. Commun*. **6**, 6962 (2015).
17. Jiang, Y. et al. Direct 3D printing of ultralight graphene oxide aerogel microlattices. *Adv. Funct. Mater*. **28**, 1707024 (2018).
18. Guo, F. et al. Highly stretchable carbon aerogels. *Nat. Commun*. **9**, 881 (2018).
19. He, P. et al. Patterned carbon nitride-based hybrid aerogel membranes via 3D printing for broadband solar wastewater remediation. *Adv. Funct. Mater*. **28**, 1801121 (2018).
20. Zhu, C. et al. Toward digitally controlled catalyst architectures: hierarchical nanoporous gold via 3D printing. *Sci. Adv*. **4**, eaas9459 (2018).
21. Chandrasekaran, S. et al. Direct ink writing of organic and carbon aerogels. *Mater. Horiz*. **5**, 1166–1175 (2018).
22. Hausmann, M. K. et al. Dynamics of cellulose nanocrystal alignment during 3D printing. *ACS Nano* **12**, 6926–6937 (2018).
23. Truby, R. L. & Lewis, J. A. Printing soft matter in three dimensions. *Nature* **540**, 371–378 (2016).
24. Gladman, A. S., Matsumoto, E. A., Nuzzo, R. G., Mahadevan, L. & Lewis, J. A. Biomimetic 4D printing. *Nat. Mater*. **15**, 413–418 (2016).
25. Kim, Y., Yuk, H., Zhao, R., Chester, S. A. & Zhao, X. Printing ferromagnetic domains for untethered fast-transforming soft materials. *Nature* **558**, 274–279 (2018).
26. Maleki, H., Montes, S., Hayati-Roodbari, N., Putz, F. & Huesing, N. Compressible, thermally insulating, and fire retardant aerogels through self-assembling silk fibroin biopolymers inside a silica structure—an approach towards 3D printing of aerogels. *ACS Appl. Mater. Interfaces* **10**, 22718–22730 (2018).
27. Hyun, K., Kim, S. H., Ahn, K. H. & Lee, S. J. Large amplitude oscillatory shear as a way to classify the complex fluids. *J. Non-Newt. Fluid Mech*. **107**, 51–65 (2002).
28. Yang, Y., Li, Y., Mao, M., Zeng, M. & Zhao, X. UV–visible–infrared light-driven thermocatalysis for environmental purification on ramsdellite MnO_2 hollow spheres considerably promoted by a novel photoactivation. *ACS Appl. Mater. Interfaces* **9**, 2350–2357 (2017).
29. He, X., Bahk, Y. K. & Wang, J. Organic dye removal by MnO_2 and Ag micromotors under various ambient conditions: the comparison between two abatement mechanisms. *Chemosphere* **184**, 601–608 (2017).
30. Wei, G., Liu, Y., Zhang, X., Yu, F. & Du, X. Thermal conductivities study on silica aerogel and its composite insulation materials. *Int. J. Heat Mass Transf*. **54**, 2355–2366 (2011).
31. Iswar, S. et al. Reinforced and superinsulating silica aerogel through in situ cross-linking with silane terminated prepolymers. *Acta Mater*. **147**, 322–328 (2018).
32. Muntz, E., Sone, Y., Aoki, K., Vargo, S. & Young, M. Performance analysis and optimization considerations for a Knudsen compressor in transitional flow. *J. Vac. Sci. Technol. A* **20**, 214–224 (2002).
33. Zhao, S. et al. Dimensional and structural control of silica aerogel membranes for miniaturized motionless gas pumps. *ACS Appl. Mater. Interfaces* **7**, 18803–18814 (2015).

Methods

Ink composition and preparation

Polyethoxydisiloxane (PEDS) precursor[34]**.** 173 ml Dynasylan 40 (ethyl silicate with an SiO_2 content of 40–42 wt%, Evonik) was mixed with 189 ml 1-pentanol (abcr Schweiz) and 13.5 ml ultrapure water (double distilled, >18 MΩ cm) at 35 °C. After stirring at 250 rpm for 10 min, the solution was cooled to 25 °C, with continuous stirring at 250 rpm. A 0.06 M HNO_3 aqueous solution was then injected dropwise (0.45 ml min^{-1}) with a syringe pump (LaboTechSystems). The silica precursor was stored at 4 °C for 24 h before use.

1-pentanol with pH indicator. 1-pentanol was selected as solvent to limit solvent evaporation during printing; 0.1 wt% methyl orange indicator (Sigma Aldrich) was added to determine the sol–gel transition.

Ramsdellite MnO_2 (R-MnO_2) microspheres[28]**.** 10.04 g $Mn(NO_3)_2{\cdot}4H_2O$ (Sigma Aldrich) was dissolved in distilled water (50 wt%). Then, 100 ml of a 3.06 wt% $KMnO_4$ aqueous solution was added dropwise and the mixture was stirred at 60 °C for 12 h. The precipitate was filtered and dried at 60 °C overnight.

Silica aerogel ink. The ink was prepared by first mixing 1-pentanol (with pH indicator) with poly(propylene glycol) bis(2-aminopropyl ether) (PPGNH; average $M_n \approx 4{,}000$; Sigma-Aldrich) at room temperature (25 °C) for 5 min. PPGNH increases viscosity and prevents settling of silica aerogel particles (Extended Data Fig. 2). Also, amino groups from PPGNH improve the homogeneity of the gel structure during the sol–gel transition (Extended Data Fig. 3). 12 M HCl (37%; Sigma-Aldrich) was added dropwise until the colour changed from yellow to red. Then, PEDS was added and mixed at 500 rpm for 5 min. Silica aerogel particles (amorphous, 5–20 μm; ENOVA, Cabot Aerogel), in mass ratios of 63–397 wt% with respect to the silica in PEDS (Extended Data Table 1), were added to achieve the rheological properties required for direct ink writing. The blend was mixed first by spatula, and then in a planetary speed-mixer (DAC 150.1 FVZ; FlackTek) at 3,000 rpm for 5 min, followed by de-foaming at 3,500 rpm for 2 min. For the R-MnO_2-silica inks, a fraction of the silica aerogel powder particles were replaced with R-MnO_2 microspheres.

Printing procedure

The silica inks were loaded into syringe barrels and de-foamed at 2,500 rpm for 3 min to remove air bubbles. The inks were then mounted to a direct ink writer from EnvisionTEC (Bioplotter) with conical nozzles (diameters of 100 μm, 250 μm, 410 μm and 1,190 μm; smoothflow tapered tip, H. Sigrist & Partner). The inks were driven pneumatically through the micronozzles at 0.3–4.0 bar, with a filament extrusion rate of 12–18.4 mm s^{-1}. Printing paths and STL files were generated by Magics Envisiontec 18.2, sliced, and converted into BPL files (Bioplotter RP software package) to command the x–y–z motion of the printer head.

Solidification and drying

The solidification of the printed objects was achieved by a base-catalysed sol–gel transition of the matrix sol (Fig. 1a). Printed objects were placed in a closed polystyrene box together with a 5.5 M ammonia solution (0.1 ml per 1 ml ink) that was not in direct contact with the objects. The NH_3 gas atmosphere induces solidification. The gelation time depends on diffusion, but typically ranges from 2 min (single-wall objects with 200–1,000 μm wall thickness) to 120 min (50 mm × 50 mm × 10 mm gel blocks) at about 25 °C. After solidification (colour change from red to yellow), ethanol (5% isopropanol, Alcosuisse) was used to cover the gels. After solvent exchange into ethanol (≥99.5%), the matrix silica gel was hydrophobized by soaking the printed objects in a 6-wt% hexamethyldisilazane (HMDZ) solution in ethanol at 25 °C for 24 h (20 ml of printed gel into 100 ml of hydrophobization solution). Finally, the objects were placed inside a SCF extractor (Separex), exchanged into liquid CO_2 over 48 h (61 bar, 23 °C), and supercritically dried from 180 bar and 48 °C.

Rheology

The rheological properties of the inks at 25 °C were characterized using a rotational rheometer (MCR502, Anton Paar) with a 50-mm-diameter steel-plate geometry and a gap height of 0.5 mm. Apparent viscosities were measured via steady-state flow experiments with a sweep of shear rate (0.001–1,000 s^{-1}). Shear storage and loss moduli were determined as a function of shear stress via oscillation experiments at a fixed frequency of 1 Hz with a sweep of stress (10–10,000 Pa). The aerogel inks were equilibrated for 1 min before testing. The shelf-life was checked after storage at 4 °C for 20 days, and the apparent viscosity and storage moduli were tested and compared with fresh ink.

Thermal conductivity

Two identical square planar boards (width, 50 mm; height, 10 mm) were printed from ink SP1.6 with a conical nozzle diameter of 1,190 μm, shown in Supplementary Video 5. After gelation and drying, the plates were placed in a custom-built guarded hot-plate device for thermal conductivity measurement (guarded zone, 50 mm × 50 mm; measuring zone, 25 mm × 25 mm; 50% relative humidity, 25 °C)[35].

Microstructural analysis

The dimension and shape of the filaments were imaged using optical (Leica DVM VZ 700C) and SEM, and the interface of the printed objects was cut using a diamond saw. SEM images were recorded on a FEI Nova NanoSEM 230 (FEI) at an accelerating voltage of 10 kV and a working distance of roughly 5 mm. Nominally 15 nm of Pt (measured with a flat piezo dector) was coated to avoid charging, but the effective thickness on the aerogels, with their extreme topography and surface area, will be much lower. For transmission electron microscopy (TEM), printed aerogel objects were crushed and suspended in methanol. The suspension was dispersed by ultrasound and dropped onto a lacey Cu TEM grid (S166-2, Plano). High-resolution TEM (HRTEM) images were recorded on a JEM-2200FS field-emission electron microscope (JEOL) at 200 kV. Nitrogen sorption analysis was carried out on a TriFlex nitrogen sorption analyser (Micromeritics) after degassing for 20 h at 100 °C and 0.03 mbar. The specific surface areas (S_{BET}; uncertainty of around 20 $m^2\,g^{-1}$) were obtained using the BET method[36]. The pore volume (V_{pore}) and average pore diameter (d_0) were calculated from the density of the printed aerogels and their S_{BET}: $V_{pore} = 1/\rho - 1/\rho_{skeleton}$ and $d_0 = 4V_{pore}/S_{BET}$, where ρ is bulk density and $\rho_{skeleton}$ is the skeletal density.

Synchrotron X-ray tomographic microscopy

A filament from ink SP1.3M0.9 printed through a 410-μm nozzle was measured by tomography. This ink was selected because the MnO_2 microparticles distributed in the matrix sol–gel provide contrast between the silica aerogel powder particles and the MnO_2-loaded silica aerogel matrix. Imaging was performed at the TOMCAT beamline of the Swiss Light Source, situated on a 2.9-T bending magnet, and equipped with a multilayer monochromator. X-ray images were acquired at 12 keV and a propagation distance of 50 mm. The X-ray indirect detector comprised a LSO:Tb 5.8-μm scintillator, a 40× optical objective and a sCMOS pco.EDGE camera (6.5 μm pixel size, 2,160 × 2,560 pixels), resulting in an effective pixel size of 0.16 μm. During the continuous tomography scan, 1,801 projections were collected over 180°, with an exposure time of 150 ms per projection, as well as two series of 100 flats and a series of 30 dark projections. The data were reconstructed using the Gridrec algorithm[37].

3D image analysis

Distinct phases are observed in the slices (Extended Data Fig. 4b), where aerogel powder grains appear darkest. The MnO_2-loaded silica aerogel matrix (binder phase) has a higher absorption and thus appears brighter. There are also distinct bright areas, probably related to MnO_2 aggregations. The size of these MnO_2-enriched areas ranges from the resolution limit of around 1 μm up to around 60 μm, but these areas are not of primary interest because they do not exist in the MnO_2-free printed filaments. A volume (203 μm × 118 μm × 234 μm) free of large MnO_2-enriched areas was selected for image analysis.

The grey values were deconvoluted in three peaks, corresponding, in order of increasing grey value, to silica aerogel particle grains, the silica aerogel matrix with finely dispersed MnO_2, and MnO_2-enriched areas (Extended Data Fig. 4a). A normal distribution of grey values within a phase was assumed, and the fit was done using the nl2sol library[38] in R[39]. The position of the grey-value distribution of the MnO_2-enriched areas was restricted to be higher than for the binder, and the weight was restricted to no more than twice the starting value, on the basis of the remainder after peak fitting of only two peaks up to a grey value of 2,000.

For the majority of the binder and the MnO_2-enriched phases, no constant contrast is reached within the phase. Instead, the grey value has a peak across the binder phase when plotted as a profile (Extended Data Fig. 4b). This indicates that most of the binder structure is at the edge of the resolution limit and the contrast gets smeared out over a larger area. To counteract this effect, the threshold grey value for the matrix aerogel (17,839) and MnO_2-enriched phase (22,860) was chosen such that, according to the phase deconvolution, the probability of a voxel (pixel) being the binder (MnO_2) phase was more than 90%. Before applying the threshold, a Gaussian blur with a width of 2 voxels or pixels was applied. After applying the threshold, the matrix aerogel and the MnO_2-enriched phase were eroded twice to remove isolated voxels or pixels and further counteract the smeared out contrast. An example of the resulting segmentation is shown in Extended Data Fig. 4b. The final segmented image consists of 57.9% SiO_2 particles, 40.5% matrix aerogel and 1.6% MnO_2-enriched areas. The final segmentation and visualization were performed using GeoDict (Math2Market).

FTIR and NMR

FTIR spectra (400–4,000 cm^{-1}) were measured on a Bruker Tensor 27 spectrometer in attenuated total reflectance mode, using a diamond crystal, and corrected for the background signal. Solid-state NMR spectra were acquired with a Bruker Avance III spectrometer equipped with a 9.4-T magnet, corresponding to Larmor frequencies of 400.2 MHz for ^1H, 100.6 MHz for ^{13}C and 79.5 MHz for ^{29}Si. Spectra were collected in 7-mm zirconia rotors with magic-angle spinning, with a spin rate of 4 kHz. ^1H–^{13}C and ^1H–^{13}Si cross-polarization spectra were collected, with contact times of 2 ms and 5 ms, respectively.

X-ray diffraction analysis

The X-ray diffraction analyses of the MnO_2 and MnO_2–SiO_2 composite were recorded on a PANalytical X'Pert PRO diffractometer equipped with a Johansson monochromator (Cu Kα1 radiation, $\lambda = 1.5406$ Å).

Thermal stability

Differential thermogravimetry analysis was conducted on a TGA7 analyser (Perkin Elmer).

Contact angle

The surface wettability by water and 1-pentanol were evaluated using a contact-angle measurement system, OCA (Dataphysics TBU 90E). Liquid droplets (5 μl) were deposited either on a packed bed of silica aerogel powder or on the surface of the printed membrane. Three measurements were performed and averaged.

Mechanical testing

Three identical cylinders (diameter, 15 mm; height, 26 mm) were prepared by casting ink SP1.6 in the cylindrical polystyrene boxes. The processed aerogel cylinders were polished to even the surfaces for uniaxial compression and Brazilian tensile testing. The specimens were tested on a mechanical testing machine (Z010, Zwick/Roell) with a 2-kN force transducer (KAP-S, AST Gruppe) at a rate of 1 mm min^{-1}. The tensile strength σ_T was calculated from the sample geometry (diameter D and height H) and Brazilian test compressive force F: $\sigma_T = F/[\pi(D/2)H]$.

Infrared thermal imaging and thermal management

The thermal insulation performance was evaluated using an infrared camera (TH 3102 MR, NEC-San-ei, Japan) equipped with a Stirling-cooled HgCdTe detector, with a temperature sensitivity of 0.08 at 30 °C and an accuracy of ±0.5 °C. The emission was set to 1. Thermal images were analysed on a PicWin-IRIS system (version 7.3). A thermal management array (Fig. 3b) was printed with a 410-μm nozzle. Practical demonstrations were carried out on a Raspberry Pi 1 Model A+ circuit board (Fig. 3f). The first demonstration was designed to illustrate anti-scalding of a heating component. As a control, the voltage-controller high-power circuit was imaged after 1 h of operation. Then, an aluminium thermal sink (70 μm thick, including conductive adhesive, sprayed with an infrared anti-reflective coating) was applied on the voltage controller to remove heat, and imaged after 1 h of operation. Finally, a 3D-printed aerogel cap (ink SP2.5, 410-μm nozzle) was placed on top and the temperature was imaged after 1 h of operation. The second demonstration was designed to protect a thermally sensitive component. For reference, a resistive heater was placed 5 mm above a capacitor (tantalum type A 476), with a T-type microthermocouple attached on top. After 30 min of stabilization, the gap distance was reduced to 1 mm for 10 s. Finally, the heater was placed directly on the capacitor and thermocouple. To benchmark the aerogel performance, a 3D-printed aerogel cap (ink SP2.5, 410-μm nozzle) and a polystyrene foam cap cut to size, both with a nominal thickness of 1.2 mm, were placed on top of the capacitor, and temperature was recorded using the same protocol.

Gas pumping and VOC degradation setup

The micropump demonstrated here relies on thermal transpiration through a porous silica aerogel membrane (Fig. 4a, b). Thermal transpiration refers to the flow of gas molecules from the cold to the hot side of a channel subjected to a temperature gradient, when the gas flow is dominated by molecular flow[40,41]. The heat is generated by the optical MnO_2 absorber layer under a 300-W halogen lamp. For thermal transpiration to be substantial, the Knudsen number $K_n = l_m/d_0$ needs to be roughly 1–10, where l_m is the mean free path of gas molecules and here d_0 is the pore chord length (which is sometimes approximated by average pore diameter d_0 above). Here l_m is calculated from the temperature (T_{avg}), pressure (P) and gas particle diameter (d_g), as $l_m = k_B T_{avg}/(\sqrt{2}\pi d_g^2 P)$, where k_B is the Boltzmann constant and $d_g = 3.711 \times 10^{-10}$ m for nitrogen gas[42]. $K_n = 2.04$ for the aerogel membrane (Fig. 4).

The thermal transpiration and VOC degradation were tested in a stainless-steel reactor (Extended Data Fig. 9). The bilayer MnO_2–SiO_2 aerogel composite was printed onto a glass fibre sheet (LydAir; 0.36 mm thickness) and the dried sample was sealed between two compartments of the reactor. The upper compartment contains a window for light irradiation on the MnO_2 absorber. The gas flow from the bottom part of the reactor to the outlets in the upper part is driven by the temperature gradient generated accross the aerogel insulator. The gas flow was monitored using a mass-flow meter (MFC, AALBORG TIO Totalizer). The VOC degradation of the functional micropump was monitored in a closed system containing synthetic air with 25 ppm toluene. First, toluene was introduced into the reactor and circulated using an external pump to reach a steady-state condition. After that, the sample was irradiated by a halogen lamp (300 W). The light intensity on the sample surface was 344 mW cm^{-2} (S170C microscope power sensor and Thorlabs PM100USB power/energy meter) and the resulting temperature increase generated a gas flow through the MnO_2-loaded catalytic layer where toluene was degraded. The concentration of toluene was monitored by gas chromatography (GC/FID). The surface temperature was monitored by a thermocouple (type K, NiCr-Ni).

Imaging and videography

Images and videos were taken with a digital SLR camera (Canon, EOS 100D) equipped with a high-performance standard zoom lens (EF-S18-55mm f/3.5-5.6 IS STM).

Data availability

The raw data on particle size distribution (Fig. 1f), rheological measurements (Fig. 1g, h), nitrogen sorption (Fig. 2m, n), thermogravimetric analysis (Fig. 2o), pumping flow rate and toluene degradation (Fig. 4f, g), and thermal conductivity measurements and reconstructed X-ray tomography used for the image analysis (Fig. 2j–l), are available at https://doi.org/10.5281/zenodo.3794969. All other data (raw data used for Extended Data Figs. 1–9 and Extended Data Table 1) are available from the corresponding authors on request.

Code availability

The codes for 3D printing and tomographic analysis are available at https://doi.org/10.5281/zenodo.3794969.

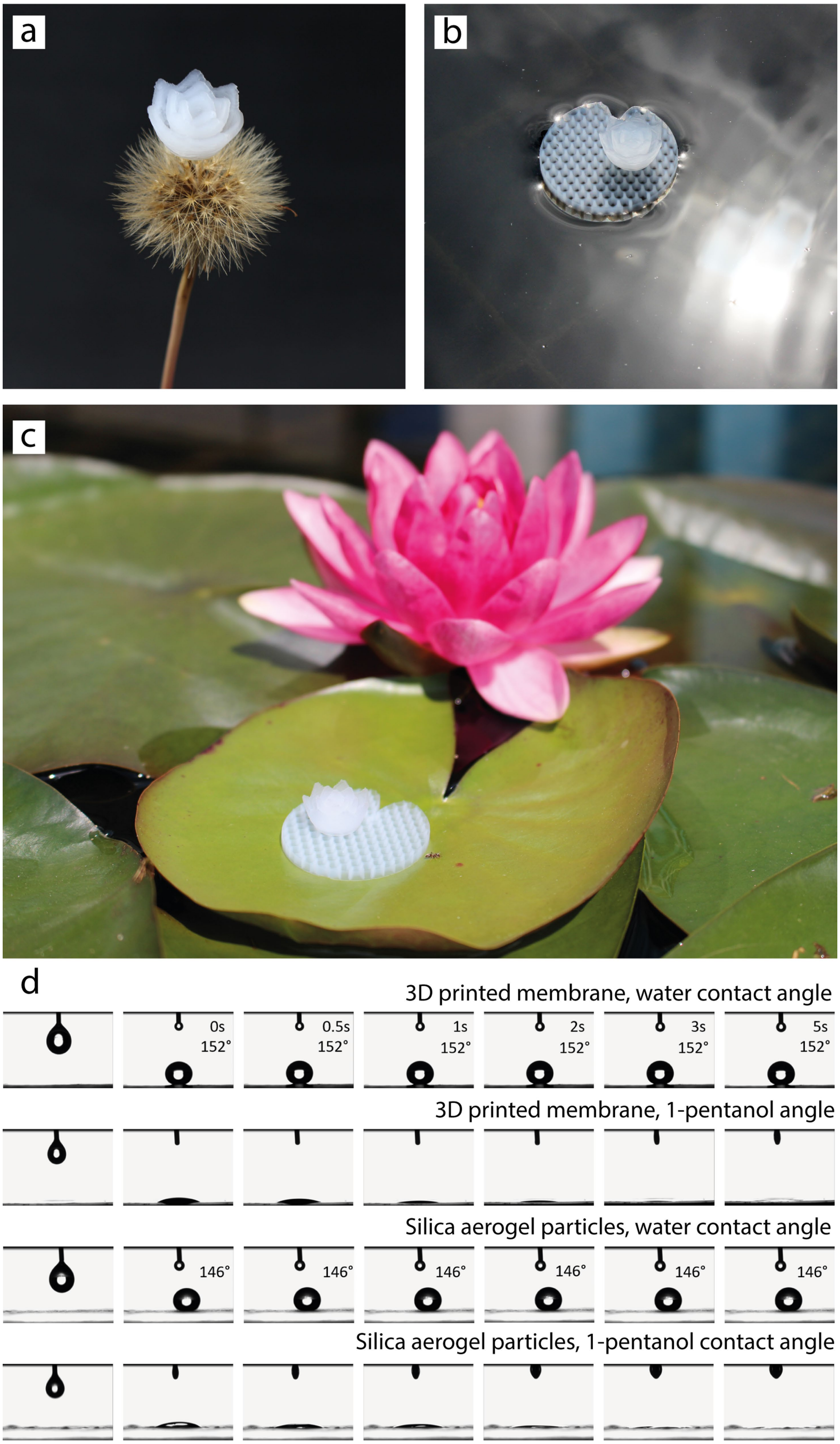
34. Pajonk, G. M. et al. Physical properties of silica gels and aerogels prepared with new polymeric precursors. *J. Non-Cryst. Solids* **186**, 1–8 (1995).
35. Stahl, T., Brunner, S., Zimmermann, M. & Ghazi Wakili, K. Thermo-hygric properties of a newly developed aerogel based insulation rendering for both exterior and interior applications. *Energy Build.* **44**, 114–117 (2012).
36. Brunauer, S., Emmett, P. H. & Teller, E. Adsorption of gases in multimolecular layers. *J. Am. Chem. Soc.* **60**, 309–319 (1938).
37. Marone, F. & Stampanoni, M. Regridding reconstruction algorithm for real-time tomographic imaging. *J. Synchrotron Radiat.* **19**, 1029–1037 (2012).
38. Fox, P., Hall, A. & Schryer, N. L. The PORT mathematical subroutine library. *ACM Trans. Math. Softw.* **4**, 104–126 (1978).
39. R: a language and environment for statistical computing (R Core Team, 2013).
40. Maxwell, J. C. VII. On stresses in rarified gases arising from inequalities of temperature. *Philos. Trans. R. Soc. Lond.* **170**, 231–256 (1879).
41. Loeb, L. B. *The Kinetic Theory of Gases* 3rd edn, Ch. VII, 290–300 (Dover, 2004).
42. Sone, Y. *Molecular Gas Dynamics Theory, Techniques, and Applications* Ch. 1 (Birkhäuser, 2007).
43. Malfait, W. J., Verel, R. & Koebel, M. M. Hydrophobization of silica aerogels: insights from quantitative solid-state NMR spectroscopy. *J. Phys. Chem. C* **118**, 25545–25554 (2014).
44. Malfait, W. J. et al. Surface chemistry of hydrophobic silica aerogels. *Chem. Mater.* **27**, 6737–6745 (2015).

Acknowledgements We thank M. Giannakou and D. Sivaraman for help in the laboratory, D. Rentsch for support with the NMR spectrometer, funded in part by grant 150638 from the Swiss National Science Foundation, F. Winnefeld for support with the rheometer, A. Braun for support with the photocatalytic setup, S. Carl for support with the infrared thermal imaging system, R. Pauer for support with the STEM and SAED pattern analysis, N. Guerrero for video editing and B. Fischer for TGA analysis. M.G. was supported by MSCA grant 746992.

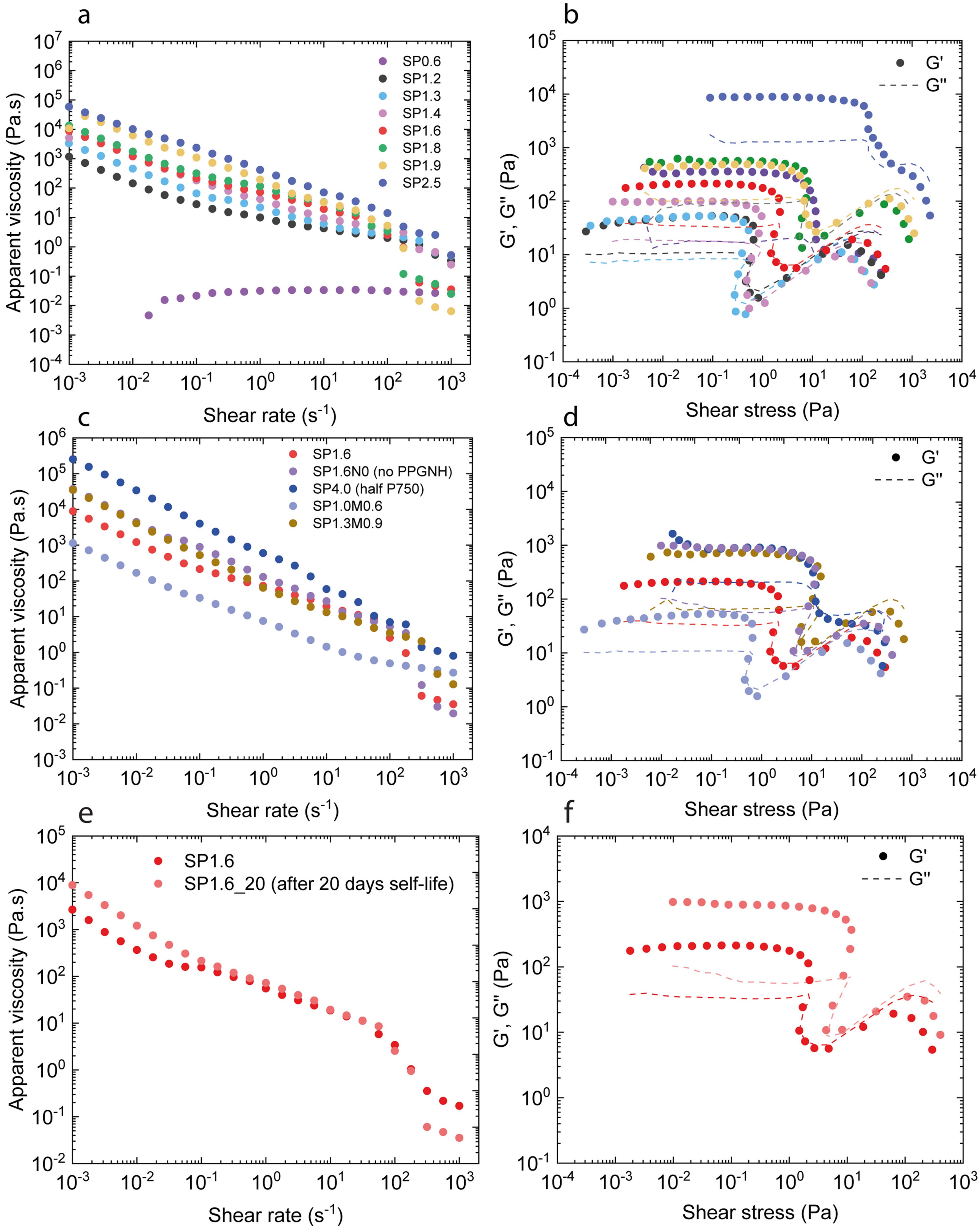
Author contributions S.Z., G.S., M.M.K. and W.J.M. designed and coordinated the study. S.Z., G.S., D.N. and C.U. optimized the ink formulation. G.S. and S.Z. carried out the printing experiments, supervised by G.N. J.W. supervised the MnO_2 synthesis and light-driven pump test. S.D. and S.Z. conducted the pumping/degradation experiments. S.Z. carried out the SEM and BET analysis. S.B. and S.Z. conducted the thermal conductivity and infrared imaging measurements. M.G. prepared STL files and photographic artwork. W.J.M. did the NMR analysis. A.B. collected, and S.G. and Z.P. processed, the X-ray tomography data. S.Z. and W.J.M. wrote the manuscript with input from all coauthors. All authors reviewed and commented on the manuscript.

Competing interests A patent for the additive manufacturing of silica aerogels (EP19208179.2) has been filed on behalf of Empa. S.Z., G.S., W.J.M., M.M.K., C.U. and D.N. are listed as inventors.

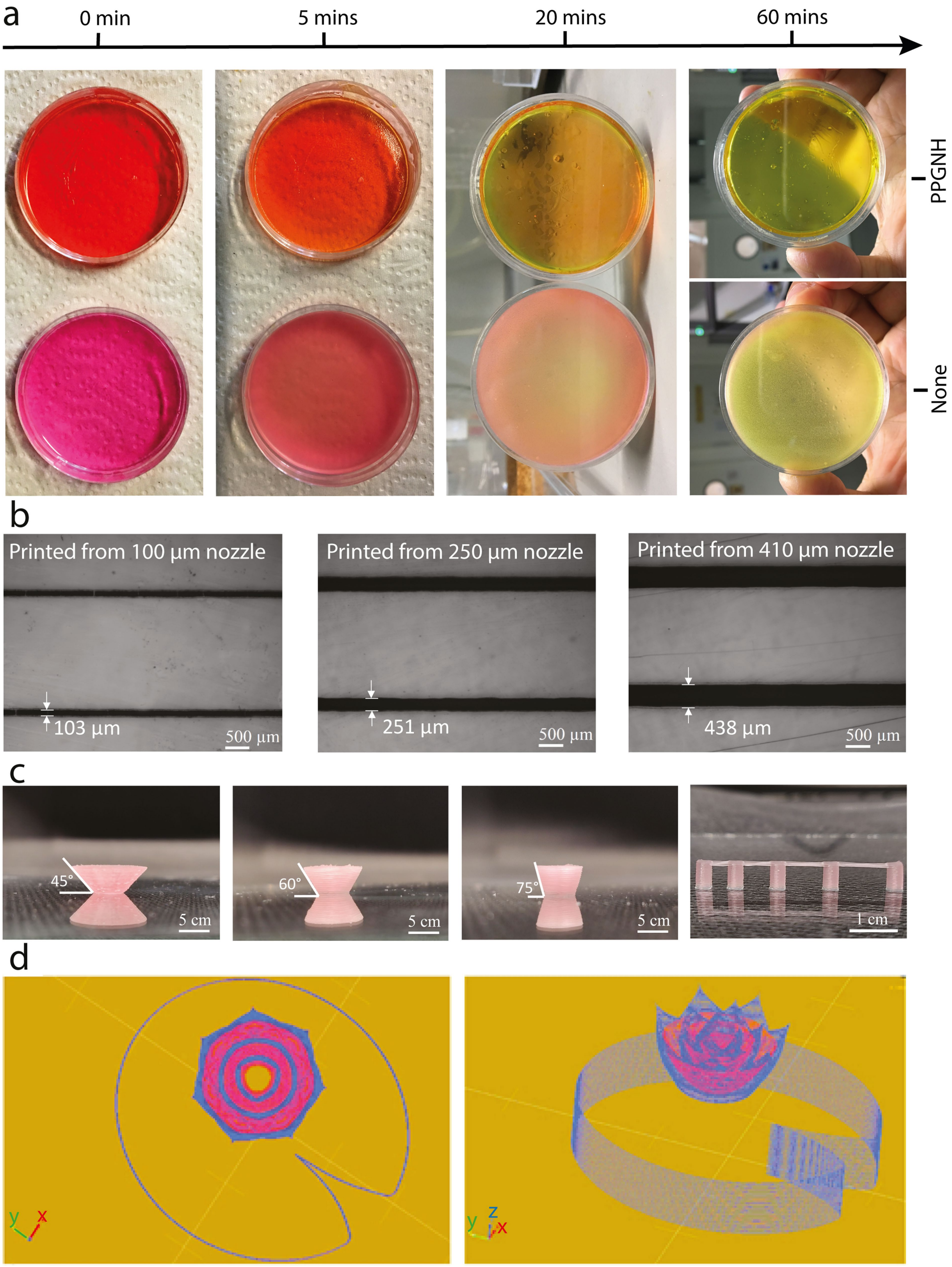
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2594-0.
Correspondence and requests for materials should be addressed to S.Z. or W.J.M.
Peer review information *Nature* thanks L. Zhao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



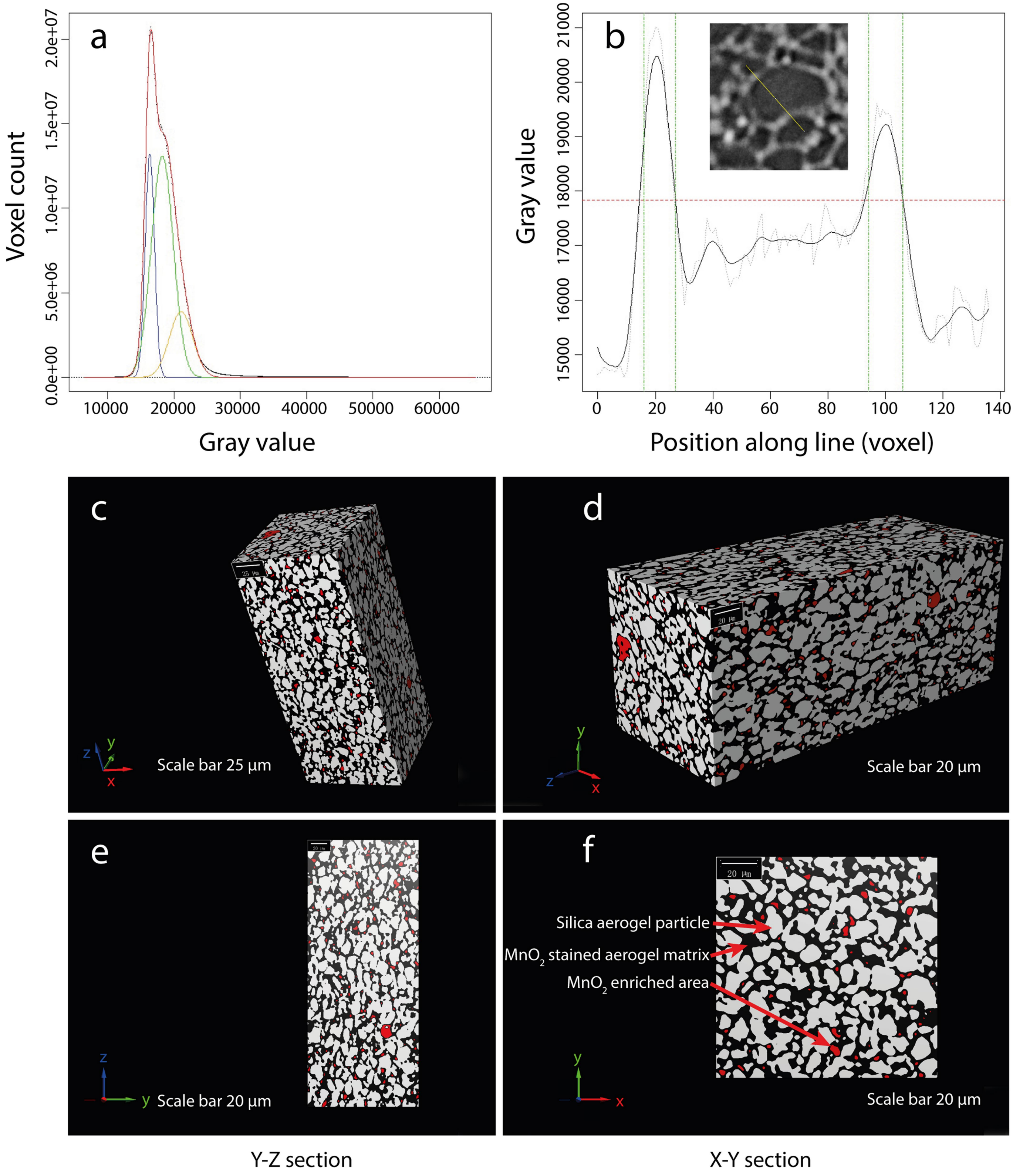
Extended Data Fig. 1 | Photographs and hydrophobicity of 3D-printed silica aerogel objects. **a**–**c**, Photographs of the 3D-printed silica aerogel lotus flower (**c**) show it is light-weight (**a**) and superhydrophobic (**b**). **d**, Water and 1-pentanol contact-angle measurements.



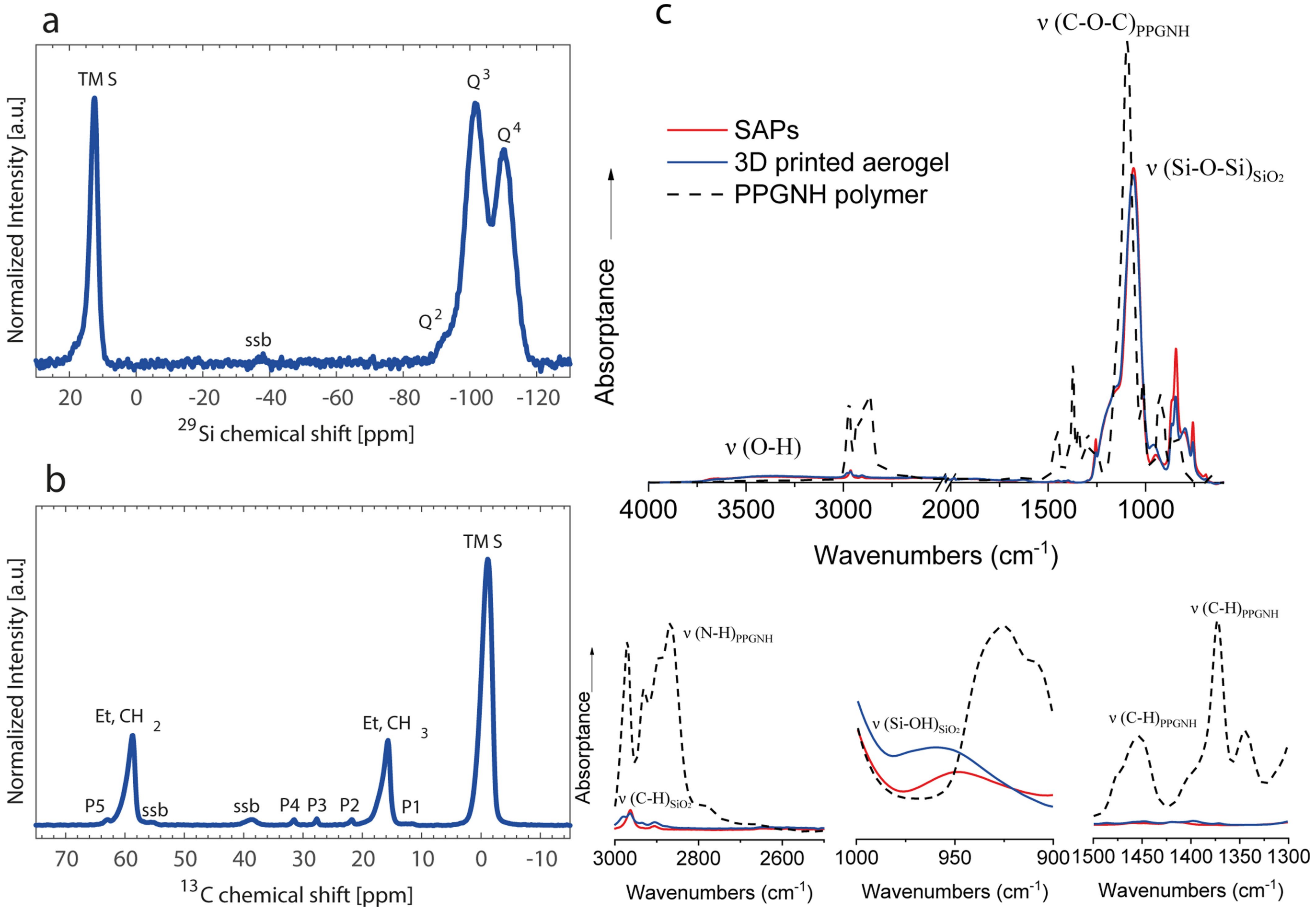
Extended Data Fig. 2 | Rheology of silica aerogel inks. a–d, Apparent viscosity as a function of applied shear rate (**a**, **c**), and storage (*G′*) and loss (*G″*) modulus as a function of applied shear stress (**b**, **d**), for the silica aerogel ink with various loadings of silica aerogel particles (as labelled). **e**, **f**, Shelf life of ink SP1.6: apparent viscosity as a function of applied shear rate (**e**), and storage (*G′*) and loss (*G″*) modulus as a function of applied shear stress (**f**), for the fresh silica aerogel ink and after 30 days of storage at room temperature (25 °C, 50% relative humidity).



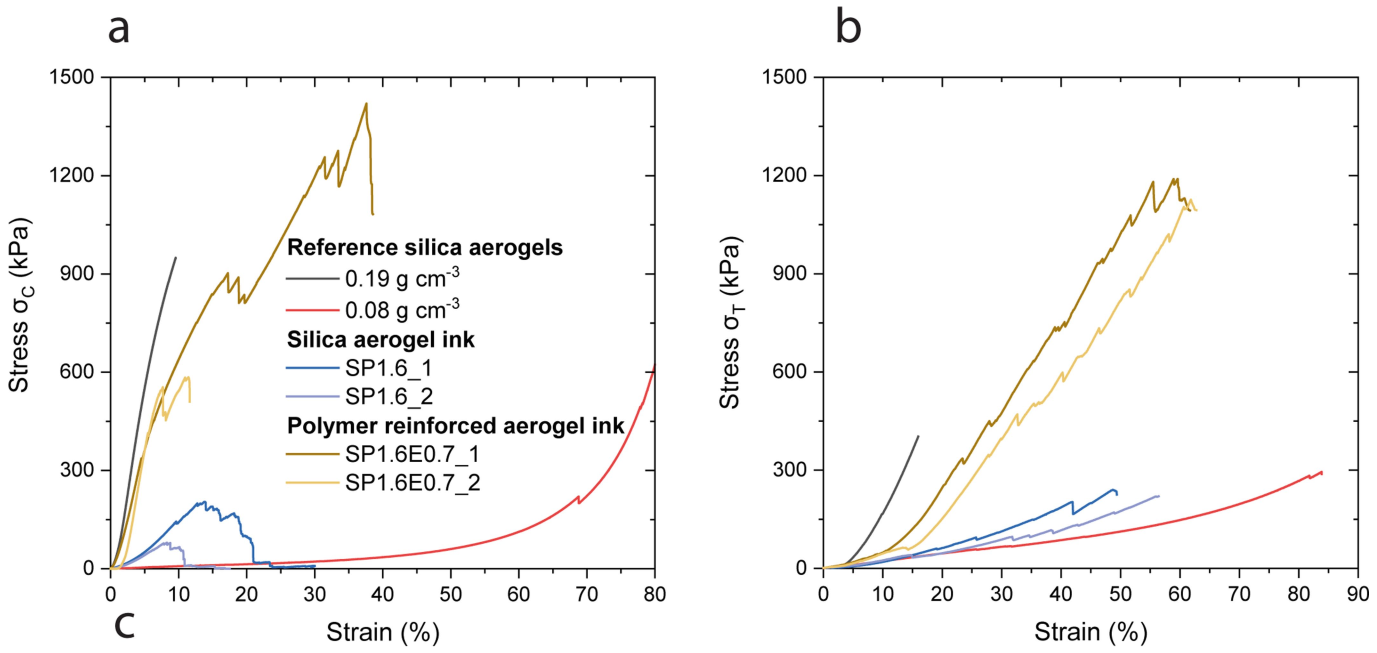
Extended Data Fig. 3 | Solidification and printing performance. a, The evolution of solidification of the silica aerogel ink with and without PPGNH. The gels are more transparent when PPGNH has been added to the sol, indicative of a more homogenous pore structure. **b**, Filaments printed from 100-µm, 250-µm and 410-µm nozzles. **c**, Demonstrations of overhang and bridging with ink SP2.5. **d**, Schematic designs of the 2D and 3D structures in Fig. 1e: a lotus flower (31–50 layers) and its leaf (24 layers). Models modified from a design by J. Watkins, https://www.thingiverse.com/thing:415314.



Extended Data Fig. 4 | X-ray tomography image analysis. a, Histogram of grey values with deconvoluted peaks of aerogel particles (blue), MnO$_2$-loaded aerogel matrix (green) and MnO$_2$-enriched areas (orange), and the resulting fit (red). **b**, Line profile across a particle (see inset), before (grey dotted line) and after (black line) Gaussian blurring. The chosen binder cutoff (red dashed line) and the final phase separation after erosion (green) are also shown. **c**–**f**, Different orientations of the 3D volume rendering of the filament (**c**, **d**), and *y*–*z* (**e**) and *x*–*y* (**f**) cross-sections. The segmented image consists of 57.9% SiO$_2$ particles, 40.5% binder phase and 1.6% MnO$_2$-enriched areas.

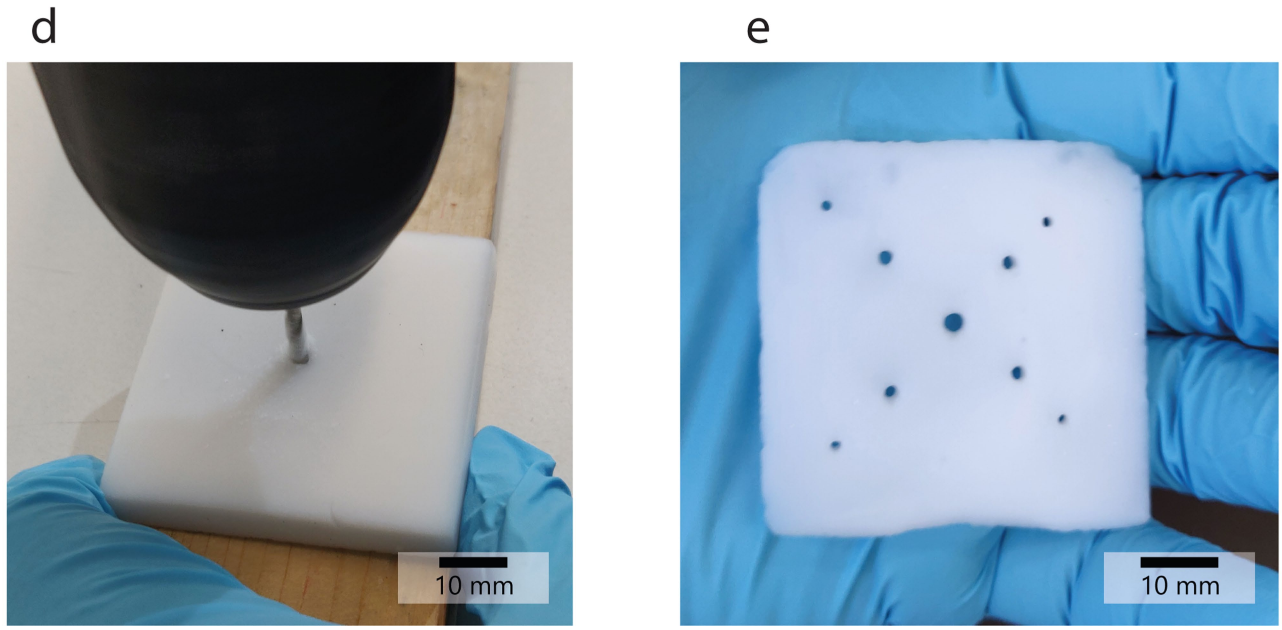


Extended Data Fig. 5 | Spectroscopic data. a, **b**, ^{1}H–^{29}Si (**a**) and ^{1}H–^{13}C (**b**) cross-polarization magic-angle spinning NMR spectra of a 3D-printed silica aerogel. The spectra have the typical peaks for hydrophobized silica aerogel[43,44]. The ^{1}H–^{29}Si spectrum has peaks Q^4, Q^3 and Q^2 from silica particles, corresponding to silicon atoms coordinated by 4, 3 and 2 bridging oxygen atoms and 0, 1 and 2 non-bridging oxygen atoms, respectively, and a peak from trimethylsilyl (TMS) groups grafted onto the silica surfaces. The ^{1}H–^{13}C spectrum has an intense peak from grafted TMS groups (63.0% of total spectral intensity) and two strong peaks from ethoxy groups grafted onto the silica surfaces (Et, CH_2 and Et, CH_3 with 34.5% of total spectral intensity), from the TEOS precursor of the matrix aerogel and/or from alcoholysis during solvent exchanges. An additional five weak peaks (P1–P5), corresponding to either pentanol adsorbed on the surface or petanoxy groups grafted onto the silica, together account for approximately 2.5% of the total spectral intensity. The peaks labelled 'ssb' correspond to spinning sidebands. No other peaks were detected. **c**, FTIR spectra of the silica aerogel particles (SAPs; red), the PPGNH (blue) and the printed object (dashed black).

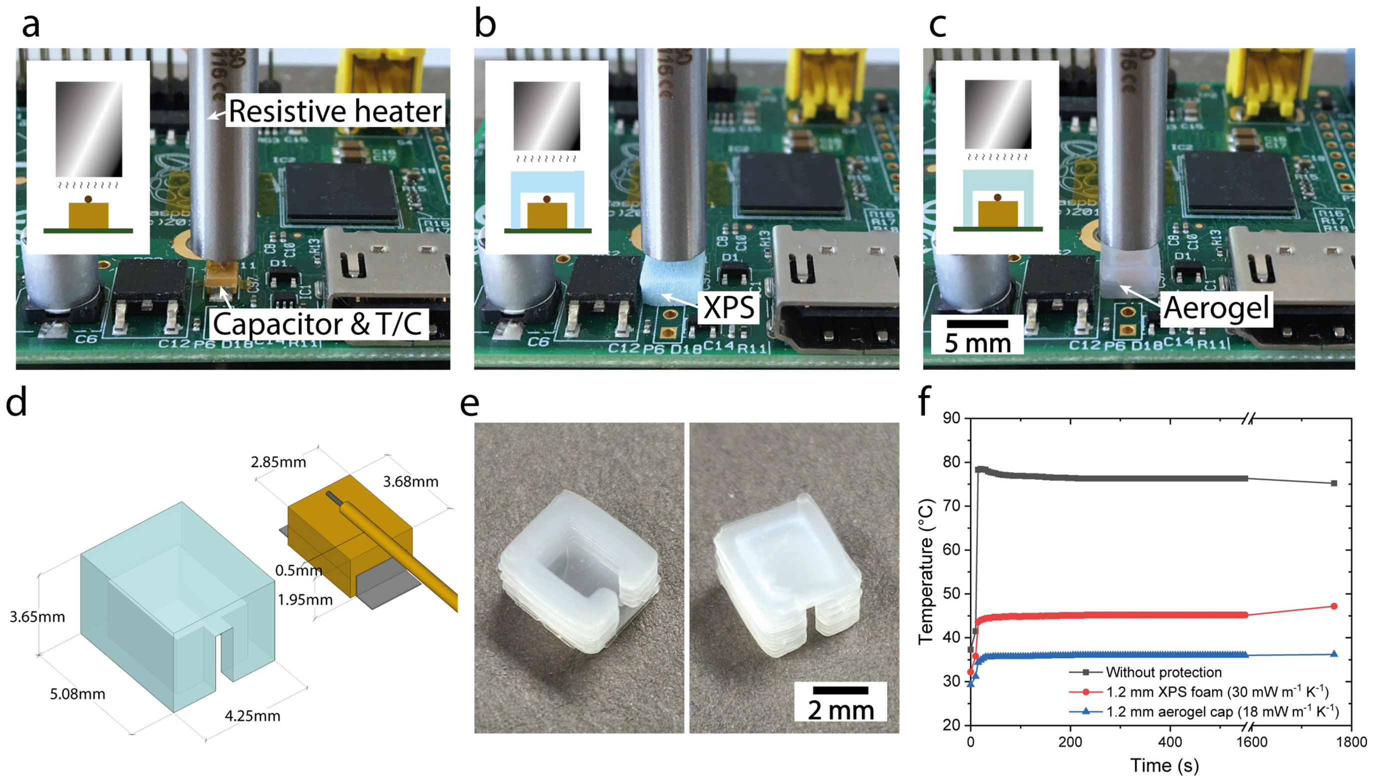


c

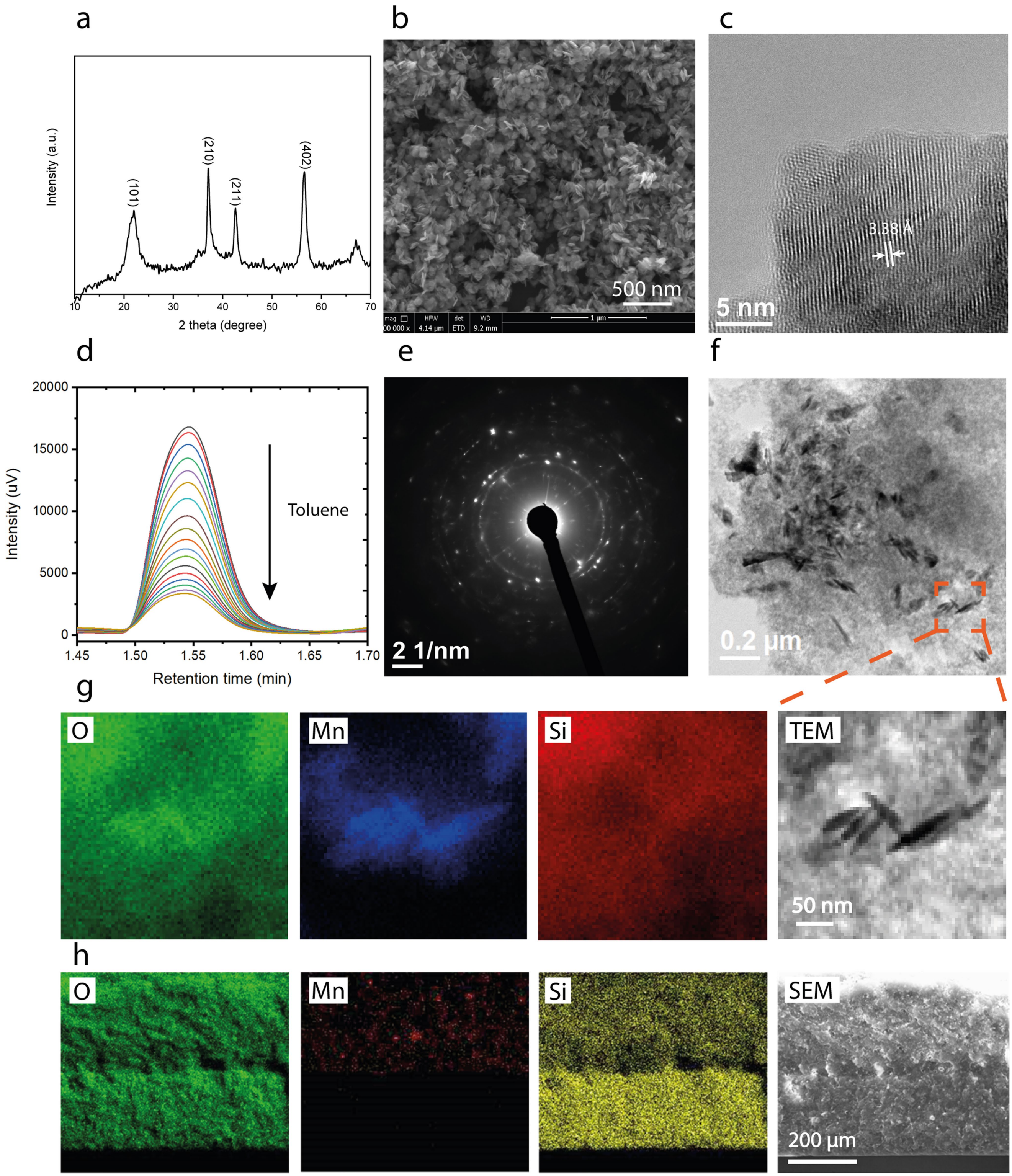
Samples	Density	σ_{Cmax}	$E_{Cmodulus}$	σ_{Tmax}	$E_{Tmodulus}$
	g cm^{-3}	MPa	MPa	MPa	MPa
Ref 0.19	0.19	0.95	13.74	0.40	1.93
Ref 0.08	0.08	0.62	0.08	0.29	0.23
Ink SP1.6 $_{ave}$	0.18	0.14	1.05	0.23	0.21
Ink SP1.6E0.7 $_{ave}$	0.28	1.00	9.6	1.14	0.53



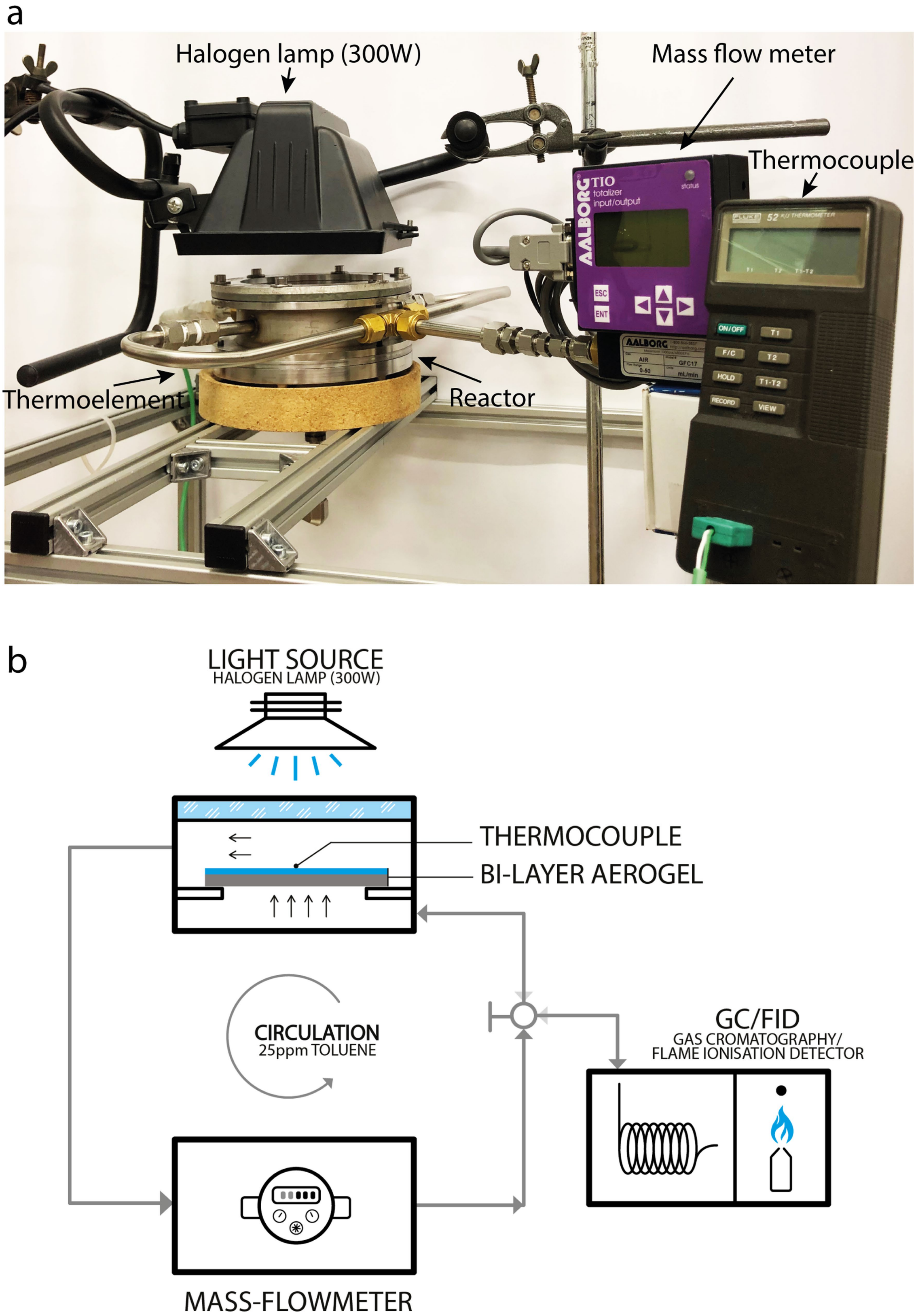
Extended Data Fig. 6 | Mechanical properties of silica-ink-based aerogels. **a**, **b**, Stress–strain curves during uniaxial compression (**a**) and Brazilian test (**b**) of reference sol–gel aerogels and silica-ink-based aerogels. **c**, The corresponding mechanical properties of the various aerogels. σ_{Cmax}, maxium compressive strength; $E_{Cmodulus}$, elastic modulus derived from the compression test; σ_{Tmax}, maxium Brazilian tensile strength; $E_{Tmodulus}$, elastic modulus derived from the Brazilian tensile strength. **d**, **e**, Subtractive manufacture of the silica-ink-based gel.



Extended Data Fig. 7 | Protection of a temperature-sensitive capacitor from a heat source. a–**c**, Without protection (T/C, T-type thermocouple; **a**), with a polystyrene foam cap (XPS; **b**) and with a printed silica aerogel cap (**c**). **d**, STL file. **e**, 3D-printed silica aerogel cap. **f**, Temperature evolution after contact with the cartridge heater.



Extended Data Fig. 8 | Properties of MnO_2 and MnO_2-doped silica aerogels. **a**–**g**, X-ray diffraction (**a**), SEM image (**b**), HRTEM image and lattice spacing (**c**), GC/FID spectra of the toluene degradation shown in Fig. 4 on a MnO_2–SiO_2 bilayer aerogel (**d**), SAED of the MnO_2 microspheres (**e**), and STEM image (**f**; **g**, right-most image) and elemental analysis (**g**, left three images) of the MnO_2 distribution in the silica aerogel. **h**, Energy-dispersive X-ray spectroscopy (right-most image) of a cross-section of the interface between the silica and MnO_2-loaded silica aerogel within the thermal transpiration membrane, and the element distribution maps (Mn, Si, O; left three images).



Extended Data Fig. 9 | Light-driven gas pump and VOC degradation system. a, **b**, Photograph (**a**) and working scheme (**b**) of the setup.

Extended Data Table 1 | Silica aerogel ink compositions and properties

Samples[a]	Compositions								Properties			
	1-Pent.	PPGNH[b]	SiO_2 in PEDS	SAPs[c]	MnO_2	EP_M95[d]	MO[e]	SiO_{2SAPs}/SiO_{2PEDS}	ρ	S_{BET}	$V_{BJHpore}$	λ
	g						mg/l	-	g cm^{-3}	m^2 g^{-1}	cm^3 g^{-1}	mW $m^{-1}K^{-1}$
SP0.6	7.3	0.7	0.8	0.5	-	-	100	0.6	-			-
SP1.2	7.3	0.7	0.8	1.0	-	-	100	1.2	0.15±0.03	622	3.0	-
SP1.3	7.3	0.7	0.8	1.05	-	-	100	1.3	-	-	-	-
SP1.4	7.3	0.7	0.8	1.14	-	-	100	1.4	0.17±0.05	794	2.5	17.4±0.2
SP1.6	7.3	0.7	0.8	1.27	-	-	100	1.6	0.18±0.02	751	3.2	15.9±0.4
SP1.6_PPGNH0	7.3	0	0.8	1.27	-	-	100	1.6	0.20±0.02	1108	3.7	17.2±0.3
SP4.0	7.3	0.7	0.32	1.27	-	-	100	4.0	-	-	-	-
SP1.8	7.3	0.7	0.8	1.41	-	-	100	1.8	-	-	-	-
SP1.9	7.3	0.7	0.8	1.52	-	-	100	1.9	0.18±0.03	748	2.6	15.5±0.3
SP2.5	7.3	0.7	0.8	2.0	-	-	100	2.5	0.22±0.02	800	2.7	18.0±0.2
SP1.0M0.6	7.3	0.7	0.8	0.8	0.5	-	-	1.0	-	-	-	-
SP1.3M0.9	7.3	0.7	0.8	1.05	0.7	-	-	1.3	-	658	3.5	17.9±0.4
SP1.6E0.7	7.3	0.7	0.8	1.27	-	2.4	100	1.6	2.8±0.01	-	-	-

ρ, bulk density; S_{BET}, surface area calculated from N_2 sorption (precision estimated to be 30 m^2 g^{-1}); $V_{BJHpore}$, average pore volume derived from the BJH analysis (precision estimated to be 0.2 cm^3 g^{-1}); *λ*, thermal conductivity; the uncertainty on *ρ* and *λ* is calculated from the standard deviation of 3–5 measurements.

[a] The sample are named using the following convention: SPxM(E)*y* denotes an ink with a mass ratio *x* of SiO_2 from silica aerogel particle ('S') to PEDS sol ('P'), and a mass ratio *y* of MnO_2 ('M') or EP_M95 prepolymer ('E') to silica in PEDS sol; SP1.6_PPGNH0 is the ink formulation without PPGNH.

[b] PPGNH, poly(propylene glycol) bis(2-aminopropyl ether).

[c] SAPs, SiO_2 aerogel particles.

[d] EP_M95 is a commercial aliphatic silane-terminated prepolymer (provided by Evonik), with a structure of $(MeO)_3$–Si–$(CH_2)_3$–(NH)–(C=O)–O–R–O–(C=O)–(NH)–$(CH_2)_3$–Si–$(OMe)_3$ and M_W of around 600 g mol^{-1}.

[e] MO, methyl orange.

Article

The causes of sea-level rise since 1900

https://doi.org/10.1038/s41586-020-2591-3

Received: 21 December 2019

Accepted: 8 June 2020

Thomas Frederikse[1]✉, Felix Landerer[1], Lambert Caron[1], Surendra Adhikari[1], David Parkes[2], Vincent W. Humphrey[3], Sönke Dangendorf[4,5], Peter Hogarth[6], Laure Zanna[7], Lijing Cheng[8,9] & Yun-Hao Wu[10]

The rate of global-mean sea-level rise since 1900 has varied over time, but the contributing factors are still poorly understood[1]. Previous assessments found that the summed contributions of ice-mass loss, terrestrial water storage and thermal expansion of the ocean could not be reconciled with observed changes in global-mean sea level, implying that changes in sea level or some contributions to those changes were poorly constrained[2,3]. Recent improvements to observational data, our understanding of the main contributing processes to sea-level change and methods for estimating the individual contributions, mean another attempt at reconciliation is warranted. Here we present a probabilistic framework to reconstruct sea level since 1900 using independent observations and their inherent uncertainties. The sum of the contributions to sea-level change from thermal expansion of the ocean, ice-mass loss and changes in terrestrial water storage is consistent with the trends and multidecadal variability in observed sea level on both global and basin scales, which we reconstruct from tide-gauge records. Ice-mass loss—predominantly from glaciers—has caused twice as much sea-level rise since 1900 as has thermal expansion. Mass loss from glaciers and the Greenland Ice Sheet explains the high rates of global sea-level rise during the 1940s, while a sharp increase in water impoundment by artificial reservoirs is the main cause of the lower-than-average rates during the 1970s. The acceleration in sea-level rise since the 1970s is caused by the combination of thermal expansion of the ocean and increased ice-mass loss from Greenland. Our results reconcile the magnitude of observed global-mean sea-level rise since 1900 with estimates based on the underlying processes, implying that no additional processes are required to explain the observed changes in sea level since 1900.

Global-mean sea level (GMSL) has increased by approximately 1.5 mm yr^{-1} (refs. [1,4,5]) over the twentieth century, modulated by large multidecadal fluctuations[6]. Changes in GMSL are the net result of many individual geophysical and climatological processes, with some of the largest contributions coming from ice-mass loss and thermal expansion of the ocean. The level of agreement between the sum of these individual contributions and the observed changes in GMSL—often described as the 'sea-level budget'—is a key indicator of our understanding of the drivers of sea-level rise[7]. Multiple studies show closure of the sea-level budget within their stated uncertainties since the 1960s and over the era of satellite altimetry since 1993[8–10]. However, rates of GMSL change and their contributions to the budget over the entire twentieth century, and especially the first half of the twentieth century, have not yet been fully explained or attributed. Previous observation-based studies concluded that the GMSL budget for the whole twentieth century could not be closed within the estimated uncertainties[2,3]. Various explanations for this non-closure have been proposed, including an overestimation of the tide-gauge-derived rates of GMSL change[11] and underestimation of the ice-sheet contribution[12], but there is no agreement yet on the cause of this discrepancy[13].

Over the past few years, revised estimates of the main known driving processes of global sea-level rise that cover the entire twentieth century have become available[14–17], the spread among different estimates of twentieth-century glacier mass loss has been reduced[18], and improved mapping methods and correction of instrumental bias have resulted in higher estimates of the contribution from thermal expansion since the 1960s[19]. In parallel, estimates of twentieth-century GMSL change have converged to lower rates than previously estimated, as a result of improved reconstruction approaches, spatial-bias correction schemes, and the inclusion of estimates of local vertical land motion (VLM) at tide-gauge locations[4,9,20]. As a result of these developments, the GMSL budget needs to be re-estimated, to determine whether the observed sea-level rise since 1900 can be reconciled with the estimated sum of contributing processes.

---

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [2]Université Catholique de Louvain, Louvain-la-Neuve, Belgium. [3]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [4]Center for Coastal Physical Oceanography, Old Dominion University, Norfolk, VA, USA. [5]University of Siegen, Siegen, Germany. [6]National Oceanography Centre, Liverpool, UK. [7]Courant Institute, New York University, New York, NY, USA. [8]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. [9]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China. [10]Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan. ✉e-mail: thomas.frederikse@jpl.nasa.gov

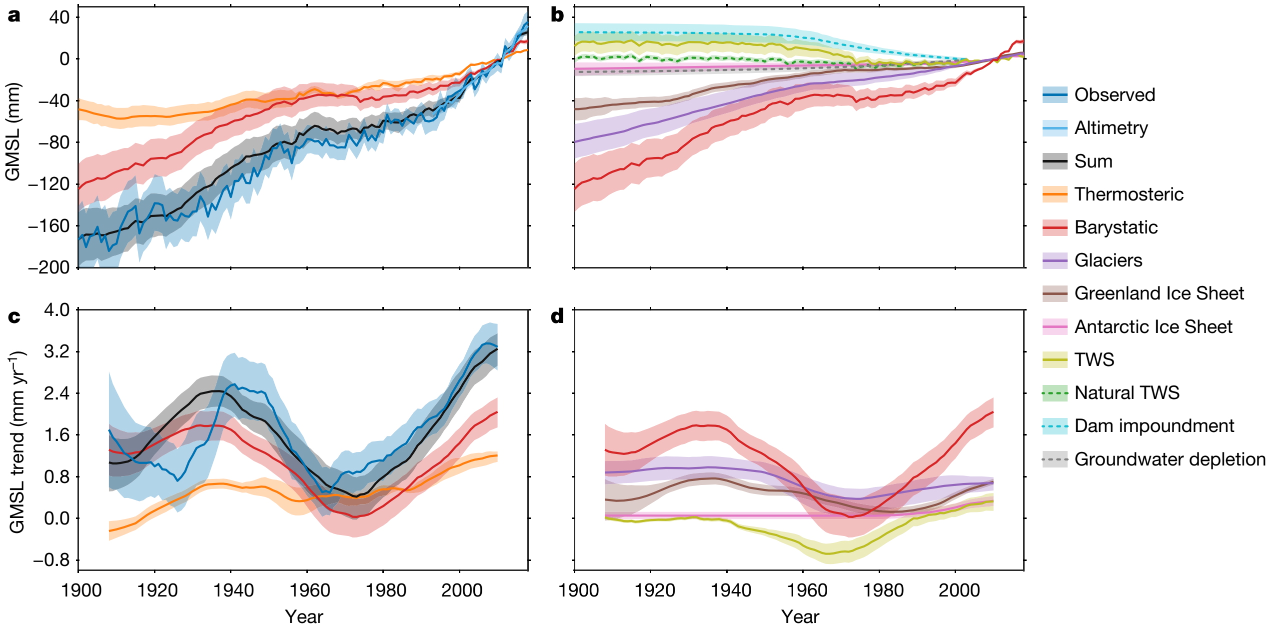


Fig. 1 | Observed GMSL and contributing processes. a, Observed GMSL, and the estimated barystatic and thermosteric contributions and their sum. **b**, The barystatic contribution and its individual components. The TWS term is the sum of groundwater depletion, water impoundment in artificial reservoirs and the natural TWS term. **c**, 30-year-average rates of observed GMSL change and of GMSL change as a result of the different contributing processes. **d**, 30-year-average rates of GMSL change due to the barystatic contribution and its individual components. The shaded regions denote 90% confidence intervals. The values in **a** and **b** are relative to the 2002–2018 mean.

Estimating the sea-level budget

To obtain estimates of changes in global ocean mass (barystatic changes), we combine estimates of mass change for glaciers[16,21], ice sheets[14,22–25] and terrestrial water storage (TWS). For the TWS estimate, we consider the effects of natural TWS variability[17], water impoundment in artificial reservoirs[26] and groundwater depletion[27,28]. For 2003–2018, we use observations from the Gravity Recovery and Climate Experiment (GRACE)[29] to quantify the barystatic changes. We estimate changes in sea level due to global thermal expansion (thermosteric changes) from in situ subsurface observations[30–32] over the period 1957–2018, and combine these estimates with an existing thermosteric reconstruction[15]. To obtain an estimate of GMSL changes and their accompanying uncertainties, we combine tide-gauge observations with estimates of local VLM from permanent Global Navigation Satellites System (GNSS) stations and with the difference between tide-gauge and satellite-altimetry observations.

Each tide-gauge and VLM record is affected by glacial isostatic adjustment (GIA) and by the effects of gravity, rotation and deformation (GRD) from contemporary surface-mass redistribution due to changes in ice mass and TWS. Owing to the irregular spatial distribution of tide-gauge sites, these effects could bias reconstructed global-mean and basin-mean sea-level changes[33]. To avoid this bias, we remove the local sea-level and VLM imprints from GIA and contemporary GRD effects from each tide-gauge and VLM record before computing basin-mean and global-mean sea-level changes from the tide gauges[9].

We propagate the uncertainties and associated covariances in the sea-level observations, in the contributing processes, and in the GIA and contemporary GRD effects into the final estimates of sea-level changes and the contributing processes. To this end, we generate an ensemble of 5,000 realizations of global-mean and basin-mean sea-level changes and all of the contributing processes. For processes for which multiple estimates are available, such as GIA, we randomly select one of these estimates when computing each individual ensemble member. For processes for which an estimate of the uncertainty is available, such as GNSS observations, we sample the estimate assuming a Gaussian distribution of the stated uncertainty about the corresponding mean. Then, we compute global-mean and basin-mean sea-level changes and the contributing processes for each ensemble member. We use the ensemble mean and spread to estimate all basin-mean and global-mean sea-level contributions and the associated confidence intervals. See Extended Data Fig. 1 and Methods for a detailed description of our approach.

Global-mean sea level

Our GMSL estimate (Fig. 1a) shows a trend of 1.56 ± 0.33 mm yr^{-1} (90% confidence interval) over 1900–2018. It is also characterized by substantial multidecadal variability, with higher rates of sea-level rise during the 1940s and since the 1990s, and lower rates around 1920 and 1970. The higher rates at the turn of the millennium are in good agreement with independent satellite-altimetry observations[34]. The observed trend over 1900–2018 is consistent with the sum of the estimated thermal expansion and changes in ocean mass, which sum to 1.52 ± 0.33 mm yr^{-1} (90% confidence interval). This consistency holds not only for the trends over the full study period, but also over the past 50 years (Table 1), and for the pattern of multidecadal variability (Fig. 1c), except for the low rates of sea-level change around the 1920s and early 1930s.

Thermosteric and barystatic sea-level changes show similar multidecadal variability patterns to the GMSL changes, although the amplitude of barystatic variability is larger than that of thermosteric variability, and barystatic variability is the main cause of multidecadal GMSL variability (Fig. 1c). The barystatic variability is not dominated by a single process (Fig. 1d). The above-average rate of GMSL rise in the 1940s is largely attributable to above-average contributions from glaciers and the Greenland Ice Sheet, whereas the high rate of barystatic sea-level rise since 2000 is attributable to both the Greenland and Antarctic ice sheets and to TWS. The low rates around 1970 are dominated by the TWS term (Fig. 1d). This negative contribution is caused predominantly by reservoir impoundment. Between 1900 and 2003, 9,400 ± 3,100 km^3 (90% confidence interval) of water has been

Table 1 | Linear trends in observed GMSL and in individual contributions to GMSL

	1900–2018 (mm yr^{-1})	1957–2018 (mm yr^{-1})	1993–2018 (mm yr^{-1})
Glaciers	0.70 [0.52, 0.89]	0.52 [0.36, 0.73]	0.67 [0.53 0.84]
Greenland Ice Sheet	0.44 [0.35, 0.53]	0.30 [0.21, 0.38]	0.65 [0.57 0.74]
Antarctic Ice Sheet	0.08 [0.00, 0.17]	0.13 [0.04, 0.22]	0.32 [0.21 0.44]
TWS	−0.21 [−0.34, −0.08]	−0.14 [−0.31, 0.02]	0.31 [0.14 0.50]
Barystatic	1.00 [0.71, 1.31]	0.80 [0.49, 1.13]	1.97 [1.63 2.33]
Thermosteric	0.52 [0.34, 0.69]	0.71 [0.54, 0.88]	1.19 [0.99 1.44]
Summed contributions	1.52 [1.20, 1.85]	1.51 [1.18, 1.84]	3.16 [2.78 3.57]
Observed GMSL	1.56 [1.24, 1.89]	1.78 [1.48, 2.07]	3.35 [2.91 3.82]
Observed GMSL minus summed contributions	0.04 [−0.31, 0.41]	0.26 [−0.07, 0.59]	0.19 [−0.32 0.70]
Satellite altimetry			3.32 [2.87 3.79]

The numbers in brackets indicate the 90% confidence interval.

impounded, leading to a sea-level drop of 26 ± 9 mm (90% confidence interval), with a peak in dam construction around the 1970s[26]. The rate of global thermosteric sea-level rise since 2000 is significantly greater than at any moment in the twentieth century. However, the barystatic rate since 2000 is not significantly greater than the rate in the 1930s. The only major feature in observed GMSL that is not replicated by the sum of the processes is the low rate in observed sea-level change during the 1920s, although this low rate is found in most ocean basins and is also visible in other reconstructions (Extended Data Fig. 2). A possible explanation for this mismatch could be the low number of available tide-gauge records over the first few decades of data, which results in a less robust reconstruction (Extended Data Fig. 3) and in increasing unquantified uncertainties in individual budget components.

The relative contributions of the barystatic and thermosteric components to GMSL vary over time. Figure 2a shows that the barystatic component dominates over the first half of the twentieth century, explaining more than 80% of total GMSL rise. The barystatic contribution is larger than the thermosteric contribution over most of the second half of the century too, except during the peak of dam construction in the 1970s. Glaciers are the largest contributor to sea-level rise over most of the twentieth century, overtaken by the thermosteric contribution only after 1970. In Fig. 2b, we omit the TWS term to remove the direct anthropogenic contributions due to reservoir impoundment and groundwater depletion. Without the TWS term, the relative contribution from glaciers and ice sheets gradually decreases during the end of the twentieth century; however, their combined contribution increases again from the start of the twenty-first century. This increase is consistent with recent assessments of the sea-level budget over the satellite era[10].

Basin-mean sea level

The global changes can be broken down into basin-mean changes (Fig. 3, Extended Data Table 1), each with different trends and variability. Although salinity-induced (halosteric) changes in sea level cause negligible changes in GMSL[35], they can be important contributors at the ocean-basin level. Thus, basin-mean changes in sea level due to changes in water density (steric changes) cannot be approximated by thermosteric changes alone[36]. Because in situ salinity estimates before the 1950s are too sparse to extract basin-scale salinity changes, we can assess the basin-mean sea-level budget only since the 1950s.

Over 1957–2018 and 1993–2018, the sea-level budget in each basin is closed within the 90% confidence intervals. The uncertainties of regional sea-level reconstructions vary considerably among basins. This is not only because of differences in tide-gauge coverage (Extended Data Fig. 3), but also to a large extent because of uncertainties in the GIA correction. In some basins, most tide-gauges are located in areas with large GIA uncertainties, such as the northwest Atlantic and the northeast Pacific coasts. On the other hand, the large uncertainties in the South Atlantic can be linked to the low number of tide-gauge records, with only a few records available before the 1960s.

In contrast to the global-mean variability, which is dominated by barystatic variability, basin-mean multidecadal sea-level variability is dominated by steric changes. The steric trends vary considerably between basins: for example, since 1957, the subtropical North Atlantic has experienced a steric trend 2.7 ± 0.4 times higher than the east Pacific. Ocean-mass trends in each basin are more homogeneous, except for the low trend in the subpolar North Atlantic. This low trend is due to the proximity to the Greenland Ice Sheet and regions of substantial glacier mass loss. Owing to GRD effects, oceans near areas of land-mass loss see below-average ocean-mass increases (Extended Data Fig. 4). This below-average increase is partially offset by GIA, which causes an upward trend in this basin. As a result, despite the fact that the observed sea-level changes in the subpolar North Atlantic can be attributed to a different mix of processes, the resulting trend since 1900 is of similar magnitude to the global-mean. GIA also results in above-average sea-level trends in the subtropical North Atlantic; for other basins, its contribution is negligible compared to ocean-mass and steric contributions. In each basin, the trend since 2000 is larger than the trend over the entire period. These high rates of GMSL change since 2000 are seen globally and are not driven by processes limited to a subset of ocean basins.

Conclusions

We reconstructed the GMSL since 1900 and compared it to the sum of the contributing processes. We found that these processes explain the

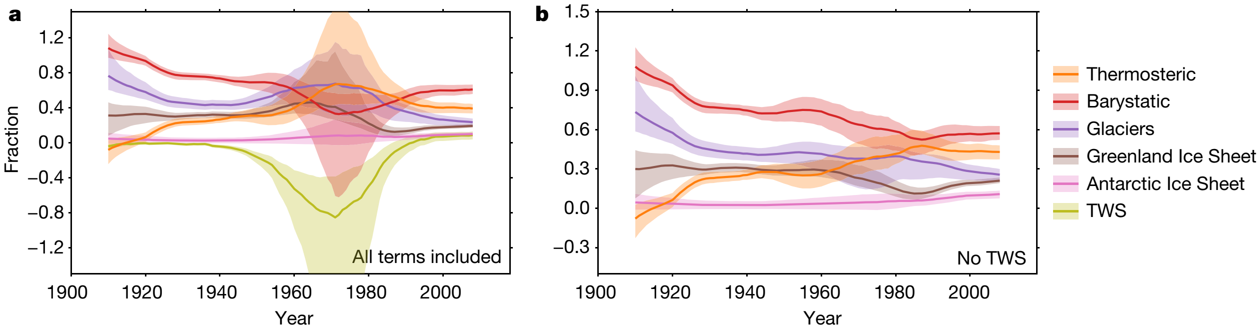


Fig. 2 | Fraction of the 40-year-average summed rate explained by each contributor. **a**, Fraction with all components included. **b**, Fraction after omitting the TWS component. The shaded regions denote 90% confidence intervals.

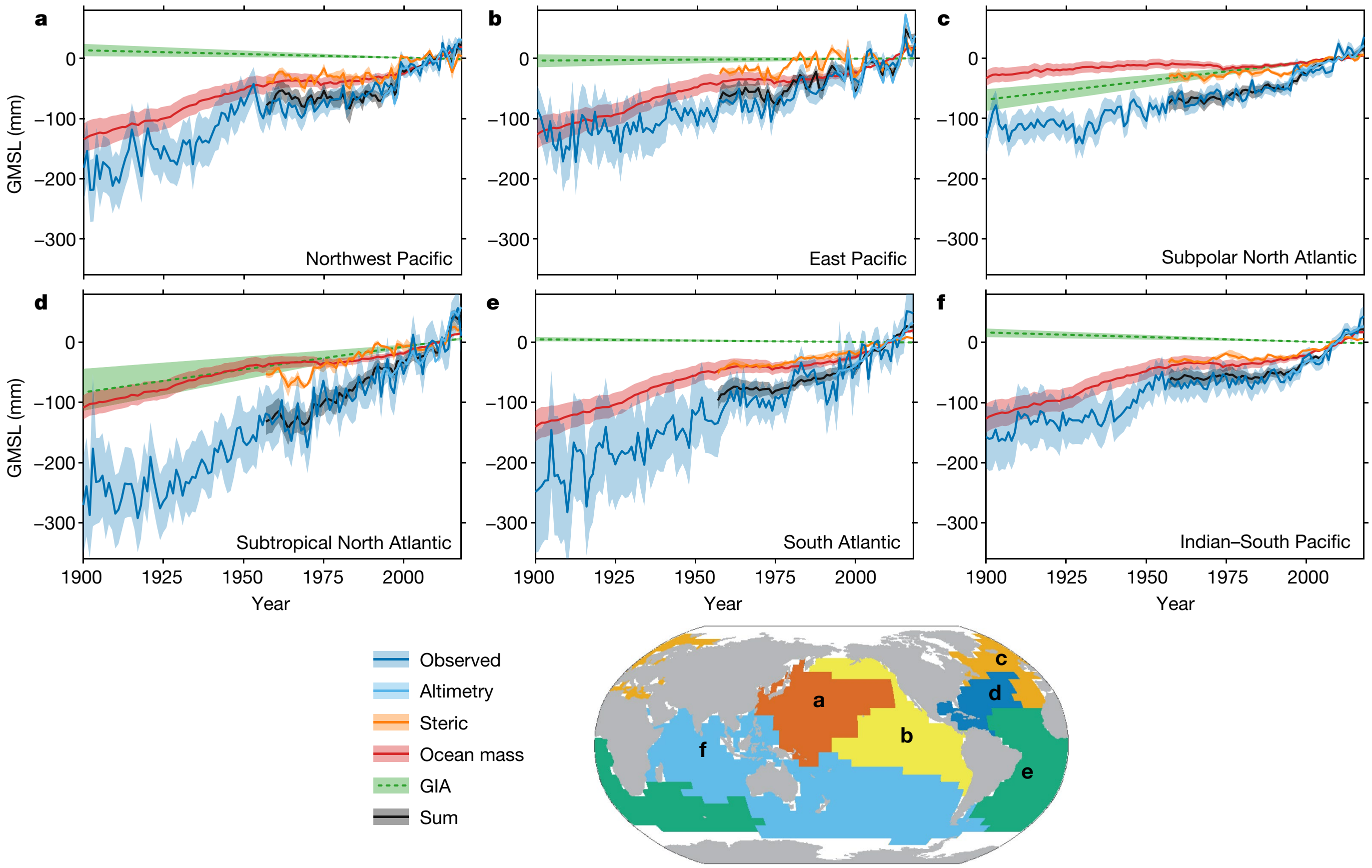


Fig. 3 | Observed basin-mean sea level and contributing processes. **a**–**f**, Observed basin-mean sea level, and the estimated contributions and their sum, for the different basins (as indicated on the map). Contrary to the global case, GIA causes basin-mean changes in sea level, and so is included in the sum of contributors. The shaded regions denote the 90% confidence interval. The values are relative to the 2002–2018 mean.

observed twentieth-century GMSL trend and match the multidecadal variability pattern, except for the low rates in observed sea-level rise during the 1920s. Barystatic changes are the primary contributor to sea-level rise, with glacier mass loss being the largest component. Reservoir impoundment caused a substantial, albeit temporary, slowdown of GMSL rise during the 1970s. The relative contributions of thermosteric and barystatic changes to GMSL vary with time. On basin scales, trends and multidecadal variability deviate from the global mean, mostly as a result of variability in the steric component.

In the subpolar North Atlantic, along which almost half of all tide gauges used in this study are located, including many of the longest available records, the ocean-mass contribution over the twentieth century is negligible, whereas GIA causes relative sea level to rise in this basin. This combination results in sea-level trends that are comparable to global-mean trends, but caused by a different combination of processes. Although many of the world's longest tide-gauge records, including the 225-year record from Amsterdam and the 220-year record from Brest, are located along the coast of the subpolar North Atlantic, long-term changes derived from these records are not representative of global-mean changes.

Closure of the twentieth-century sea-level budget, as demonstrated here, implies that no additional unknown processes, such as large-scale deep-ocean thermal expansion or additional mass loss from the Antarctic Ice Sheet, are required to explain the observed changes in global sea level. Such additional processes had been speculated to explain the non-closure found in previous studies of global sea-level budget[2,3,12]. Our demonstration of closure of the global-mean and basin-mean sea-level budget forms a consistent baseline against which process-based and semi-empirical sea-level projections can be benchmarked, without the need to compare against either the sum of processes or observed sea level[37]. The downward revision of the estimated sea-level rise and updated estimates of the driving processes, particularly the increased estimated glacier mass loss, result in a consistent picture of twentieth-century GMSL rise and its underlying causes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2591-3.

1. Oppenheimer, M. et al. in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (eds Pörtner, H.-O. et al.) Ch.4 (in the press).
2. Moore, J., Jevrejeva, S. & Grinsted, A. The historical global sea-level budget. *Ann. Glaciol.* **52**, 8–14 (2011).
3. Gregory, J. M. et al. Twentieth-century global-mean sea level rise: is the whole greater than the sum of the parts? *J. Clim.* **26**, 4476–4499 (2013).
4. Hay, C. C., Morrow, E., Kopp, R. E. & Mitrovica, J. X. Probabilistic reanalysis of twentieth-century sea-level rise. *Nature* **517**, 481–484 (2015); erratum **552**, 278 (2017).
5. Dangendorf, S. et al. Persistent acceleration in global sea-level rise since the 1960s. *Nat. Clim. Change* **9**, 705–710 (2019).
6. Chambers, D. P., Merrifield, M. A. & Nerem, R. S. Is there a 60-year oscillation in global mean sea level? *Geophys. Res. Lett.* **39**, L18607 (2012).
7. Munk, W. Twentieth century sea level: an enigma. *Proc. Natl Acad. Sci. USA* **99**, 6550–6555 (2002).
8. Church, J. A. et al. Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.* **38**, L18601 (2011); erratum **40**, 4066 (2013).
9. Frederikse, T., Jevrejeva, S., Riva, R. E. M. & Dangendorf, S. A consistent sea-level reconstruction and its budget on basin and global scales over 1958–2014. *J. Clim.* **31**, 1267–1280 (2018).

10. WCRP Global Sea Level Budget Group. Global sea-level budget 1993–present. *Earth Syst. Sci. Data* **10**, 1551–1590 (2018).
11. Cabanes, C. Sea level rise during past 40 years determined from satellite and in situ observations. *Science* **294**, 840–842 (2001).
12. Miller, L. & Douglas, B. C. Mass and volume contributions to twentieth-century global sea level rise. *Nature* **428**, 406–409 (2004).
13. Jevrejeva, S., Matthews, A. & Slangen, A. The twentieth-century sea level budget: recent progress and challenges. *Surv. Geophys*. **38**, 295–307 (2017).
14. Kjeldsen, K. K. et al. Spatial and temporal distribution of mass loss from the Greenland Ice Sheet since ad 1900. *Nature* **528**, 396–400 (2015).
15. Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J. & Heimbach, P. Global reconstruction of historical ocean heat storage and transport. *Proc. Natl Acad. Sci. USA* **116**, 1126–1131 (2019).
16. Parkes, D. & Marzeion, B. Twentieth-century contribution to sea-level rise from uncharted glaciers. *Nature* **563**, 551–554 (2018).
17. Humphrey, V. & Gudmundsson, L. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth Syst. Sci. Data* **11**, 1153–1170 (2019).
18. Marzeion, B., Leclercq, P. W., Cogley, J. G. & Jarosch, A. H. Brief Communication: Global reconstructions of glacier mass change during the 20th century are consistent. *Cryosphere* **9**, 2399–2404 (2015).
19. Cheng, L., Abraham, J., Hausfather, Z. & Trenberth, K. E. How fast are the oceans warming? *Science* **363**, 128–129 (2019).
20. Dangendorf, S. et al. Reassessment of 20th century global mean sea level rise. *Proc. Natl Acad. Sci. USA* **114**, 5946–5951 (2017).
21. Zemp, M. et al. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**, 382–386 (2019); erratum **577**, E9 (2020).
22. Adhikari, S. et al. What drives 20th century polar motion? *Earth Planet. Sci. Lett*. **502**, 126–132 (2018).
23. The IMBIE team. Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature* **558**, 219–222 (2018).
24. Bamber, J. L., Westaway, R. M., Marzeion, B. & Wouters, B. The land ice contribution to sea level during the satellite era. *Environ. Res. Lett*. **13**, 063008 (2018); corrigendum **13**, 099502 (2018).
25. Mouginot, J. et al. Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. *Proc. Natl Acad. Sci. USA* **116**, 9239–9244 (2019).
26. Chao, B. F., Wu, Y. H. & Li, Y. S. Impact of artificial reservoir water impoundment on global sea level. *Science* **320**, 212–214 (2008).
27. Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T. & Eicker, A. Global-scale assessment of groundwater depletion and related groundwater abstractions: combining hydrological modeling with information from well observations and GRACE satellites. *Wat. Resour. Res*. **50**, 5698–5720 (2014).
28. Wada, Y. et al. Fate of water pumped from underground and contributions to sea-level rise. *Nat. Clim. Change* **6**, 777–780 (2016).
29. Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C. & Landerer, F. W. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth* **120**, 2648–2671 (2015).
30. Levitus, S. et al. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett*. **39**, L10603 (2012).
31. Ishii, M. et al. Accuracy of global upper ocean heat content estimation expected from present observational data sets. *Sci. Online Lett. Atmos*. **13**, 163–167 (2017).
32. Cheng, L. & Zhu, J. Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations. *J. Clim*. **29**, 5393–5416 (2016).
33. Thompson, P. R., Hamlington, B. D., Landerer, F. W. & Adhikari, S. Are long tide gauge records in the wrong place to measure global mean sea level rise? *Geophys. Res. Lett*. **43**, 10403–10411 (2016).
34. Beckley, B. D., Callahan, P. S., Hancock, D. W., Mitchum, G. T. & Ray, R. D. On the "cal-mode" correction to TOPEX satellite altimetry and its effect on the global mean sea level time series. *J. Geophys. Res. Oceans* **122**, 8371–8384 (2017).
35. Gregory, J. M. et al. Concepts and terminology for sea level: mean, variability and change, both local and global. *Surv. Geophys*. **40**, 1251–1289 (2019).
36. Durack, P. J., Wijffels, S. E. & Gleckler, P. J. Long-term sea-level change revisited: the role of salinity. *Environ. Res. Lett*. **9**, 114017 (2014).
37. Mengel, M. et al. Future sea level rise constrained by observations and long-term commitment. *Proc. Natl Acad. Sci. USA* **113**, 2597–2602 (2016).

Methods

The global-mean and basin-mean sea-level changes that we report are relative sea-level (RSL) changes[35], corresponding to the total change in sea-water volume. RSL changes are changes relative to the underlying seafloor. They differ from geocentric sea-level changes observed by satellite altimetry, owing to seafloor deformation. We divide the global ocean into six basins[38]. These basins (Extended Data Fig. 3) are defined using a clustering approach that merges locations that share a common interannual sea-level variability signal, as observed by satellite altimetry. We define the global ocean as the sum of all basins. Our basins do not cover the highest latitudes of polar oceans, as satellites cannot sufficiently provide data for these regions. Sea-level changes in these regions, which cover 7% of the total ocean area, are not included. Because the omitted area is small, only a large local anomaly in sea-level rise would have to potential to affect GMSL substantially. A recent sea-level reconstruction[5] estimates a rate of sea-level rise of 1.0 ± 0.8 mm yr^{-1} in the Arctic ocean and a rate of 1.6 ± 0.6 mm yr^{-1} in the Southern Ocean over 1900–2015. Using these rates to extend our reconstruction has a negligible (less than 0.1 mm yr^{-1}) effect on the global-mean sea-level trend. Therefore, omitting these oceans when reconstructing global-mean sea-level changes is unlikely to cause substantial GMSL changes.

The ensemble approach

Assessing closure of the global-mean and basin-mean sea-level budget requires an estimate of the mean and associated uncertainties of the observed sea-level changes, as well as those of the major contributing processes. Some processes, especially GIA, affect both the sea-level observations and estimates of the contributing processes, and the reconstructed sea-level changes and the sum of processes are not fully independent. Therefore, we use a Monte Carlo approach to obtain a consistent set of observed sea level, its contributing processes and associated uncertainties. We generate 5,000 realizations of observed sea level and the contributing processes. For each process, we use one of the two following approaches. If a large number of estimates is available, we randomly select one estimate (for example, GIA). If only a single or limited number of independent estimates are available (for example, glacier mass loss), we generate ensemble members by randomly selecting and perturbing one of these estimates. We perturb the estimate by drawing random numbers from a Gaussian distribution using the a priori uncertainty of that estimate as the standard deviation and adding these random numbers to the estimate. We compute basin-mean and global-mean sea-level changes and the contributing processes for each ensemble member. This procedure provides 5,000 realizations of global-mean and basin-mean sea level, all components, and the difference between sea level and the sum of the components, in which all known sources of uncertainty and the spread among different estimates have been propagated. We compute all the time series, moving trends and linear trends for each ensemble member and subsequently derive the mean and confidence intervals from the ensemble. This procedure ensures that the underlying co-variances between the sea-level observations and contributing processes are propagated into the final estimates. Extended Data Fig. 1 shows the procedure that is followed for each individual ensemble member. In the sections below, we describe the data and estimates used for reconstructing sea level and each process.

GIA

While not changing contemporary GMSL, GIA causes changes in the Earth's gravity field and the shape of the solid Earth, and changes local relative sea level. These changes affect observations from tide gauges, altimetry, GNSS stations and our estimates of the contributors to barystatic sea-level changes[39]. Estimates of GIA-induced changes in sea level, gravity and the solid Earth all come with a substantial uncertainty. Because GIA input parameters simultaneously affect several components of the sea-level budget, these components and their uncertainties are not fully independent of each other[39,40]. To estimate the GIA effects and to propagate the mutually dependent uncertainties in the GIA predictions into all affected observations, we use an ensemble of GIA estimates[41]. This study[41] provides a 128,000-member ensemble of GIA predictions, computed by varying solid-Earth parameters (lithosphere thickness and mantle viscosities) and amplitudes of global deglaciation histories over the past 20,000 years. Each GIA ensemble member provides a consistent set of changes in relative sea level, solid-Earth deformation and changes in equivalent water height, used to correct GRACE observations, and comes with a likelihood that reflects how good the fit is to a dataset of vertical GNSS velocities and palaeo sea-level records. Therefore, this model allows for a robust quantification of the uncertainties associated with GIA. The spread between the ensemble members depicts the uncertainty in the GIA predictions due to uncertainty in the solid-Earth parameters and the deglaciation history. Large uncertainties can therefore be found around the edges of formerly glaciated regions, such as the coastlines of Alaska and Fennoscandia, and the forebulge collapse regions along the North American coastlines. The ensemble approach ensures that these uncertainties are propagated into estimates of basin-mean and global-mean sea level. See ref. [41] for further details about the GIA predictions and the data used to weigh the GIA ensemble members. For each of our ensemble members, we randomly select one GIA prediction from the 128,000-member ensemble. Extended Data Fig. 4b shows the ensemble-mean RSL changes caused by GIA. Using the ICE6G D (VM5a) model[42] to account for GIA (Extended Data Fig. 5) does not cause noteworthy differences in global-mean and basin-mean observed sea level and the contributing processes. The differences in the subtropical North Atlantic basin are slightly larger (up to 0.3 mm yr^{-1}), but even here the GIA-related sea-level changes are within the confidence intervals of our GIA ensemble.

Contemporary mass redistribution

For the sea-level changes due to contemporary mass redistribution, we need to estimate the amount of water that is redistributed, and where on land the water is added or removed. During 2003–2018, we use GRACE and GRACE-FO observations, based on the JPL RL06 mascon solution[29,43,44]. This solution provides monthly land-mass changes on a nominal 3-degree grid, from which we compute annual averages. Each grid cell has an associated measurement uncertainty, based on the formal error covariance matrix of the GRACE solution[43]. For each ensemble member, we randomly draw from these uncertainty estimates, perturb the mass estimates with this draw and correct for GIA. We then split the land-mass changes from GRACE into mass changes from glaciers, ice sheets and TWS using a previously described method[45].

Over 1900–2003, we use multiple estimates of each of the aforementioned processes. To combine these estimates with the GRACE observations, we average all observation-based mass-loss estimates over the same grid as the GRACE observations and remove the common mean in 2003 at every GRACE grid cell. Extended Data Fig. 6 shows all individual estimates and the resulting final composite estimate for each mass-redistribution process.

For glaciers, we use two mass-change estimates. The first estimate, which covers the whole twentieth century, is based on a global glacier model that is driven by observation-based surface forcing[18]. This model produces estimates of the annual rate of glacier mass loss for each of the 19 glaciated regions defined in the Randolph Glacier Inventory (RGI)[46]. The second estimate[21], which provides mass changes since 1961, uses in situ glaciological and geodetic observations to derive total mass changes for each glaciated region. Both estimates provide uncertainties of the rate. For each ensemble member, we randomly choose between the two estimates. Before 1961, each member uses the estimates from the first estimate. Both estimates provide annual rate uncertainties. We draw random numbers using the rate uncertainties as the standard deviation, add them to the estimated rate and integrate this perturbed rate to obtain the total glacier mass changes. Because GRACE cannot distinguish the contributions from the Greenland and Antarctic peripheral glaciers from those from the ice sheets, we do not include these glaciers into the glacier mass balance. For Greenland, we add the peripheral glaciers to the ice-sheet contribution. For Antarctica, the mass balance of its peripheral glaciers is very uncertain, owing to the lack of observations[47]. However, since 2003, only a very small mass loss has been observed for these glaciers[48], and observations since the 1950s do not suggest a large contribution[21]. Therefore, we assume no mass loss from the Antarctic peripheral glaciers. We account for missing (owing to their relatively small size) and disappeared glaciers using a previous estimate[16]. This study[16] provides upper- and lower-bound estimates of the contribution of missing and disappeared glaciers. For each ensemble member, we uniformly sample between the upper- and lower-bound estimates. Since this estimate does not provide glacier mass changes per RGI region, we assume that the regional distribution of the contribution from missing and disappearing glaciers can be scaled by the regional relative contribution from the large glaciers as recognized by RGI.

For the Greenland Ice Sheet, we use three estimates: a mass-balance reconstruction[14] that covers 1900–2003, input–output estimates[25] that cover 1972–2003, and a multi-method assessment[24] that covers 1993–2003. For each ensemble member, we randomly select one of these models. We use the first estimate for the contribution over the era for which the others do not provide an estimate. Each estimate provides a rate uncertainty, and we use these uncertainties to generate a perturbed estimate for each ensemble member using the same procedure as for glaciers. These reconstructions (except the one from ref. [24]) do not include the contribution from peripheral glaciers. For these estimates, we add the estimated peripheral glacier contribution to the Greenland mass balance using the same approach as for other glaciated regions.

For Antarctica, no mass-balance reconstruction exists before the satellite era, although observational evidence suggests twentieth-century mass loss, especially from West Antarctica[49,50]. Therefore, we assume a small Antarctic Ice Sheet contribution before 1993 of 0.05 ± 0.04 mm yr^{-1}, based on an existing compilation[22]. For 1993–2003, we use the multi-method assessments[23,24] to derive the mass changes. To obtain an estimate of the spatial pattern of the mass changes from both ice sheets, we derive the spatial pattern of the mass loss from the perturbed GRACE observations. We assume this spatial pattern remains constant in time.

The TWS component consists of natural and anthropogenic processes. For natural TWS, we use a twentieth-century reconstruction[17] that provides 100 ensemble members of natural TWS changes. We mask out all glacier and ice-sheet regions from these estimates, and randomly select one of the 100 TWS ensemble members. For anthropogenic TWS changes, we consider artificial reservoir impoundment and groundwater depletion. For reservoir impoundment, we use an updated list of global artificial reservoirs[26] and the ICOLDS dam database[51]. We assume the filling and seepage rates of each reservoir follow previous estimates[26]. The ICOLDS dam database, which covers 93% of the total impounded volume, provides location coordinates of each reservoir; the database from ref. [26] does not. To approximate the regional distribution of this reservoir impoundment, we add the impounded water of the reservoirs with unknown location to the reservoirs with known location. We compute the fraction of the total impounded volume held by each known reservoir, and distribute the water from reservoirs with unknown location using this fraction. To our knowledge, for reservoir impoundment, no formal uncertainties have been quantified. Likely sources of the uncertainty in the reservoir impoundment stem from reservoir filling levels, storage-capacity loss due to sedimentation and seepage effects[3,52]. Previous assessments assumed rates of uncertainties of 10%–30%[8,28]; we assume an uncertainty of 20% (1σ).

For groundwater depletion, we use two gridded depletion estimates. Ref. [53] provides depletion estimates over 1900–2003. However, a substantial fraction of the depleted groundwater remains on land rather than ending up in the ocean[28]. To account for this effect, we assume that 40% of the depleted groundwater stays on land, and we scale the estimated depletion from this study by a factor of 0.6 (ref. [54]). We also use depletion estimates[27] over 1961–2003. Similarly to the glacier and ice-sheet case, we randomly select one of the estimates for each ensemble member. We assume an uncertainty of 20% (1σ) in groundwater depletion, which corresponds to previously estimated uncertainties[28].

These land-mass changes result in barystatic sea-level changes and, owing to GRD effects, in regionally varying sea-level change and solid-Earth deformation patterns. For each ensemble member, we solve the sea-level equation using a pseudo-spectral method[55,56]. The spherical-harmonics transformations are computed using the SHTns library[57] up to degree and order 360. The resulting geoid changes and deformation are expressed relative to the centre-of-mass reference frame, and include rotational feedback[58]. We assume an elastic solid-Earth response to the land-mass changes, for which we use Love numbers based on[59] the Preliminary Referenced Earth Model[60]. With this procedure, we obtain 5,000 ensemble members, each consisting of annual time series of local sea-level changes and solid-Earth deformation due to contemporary mass redistribution. Extended Data Fig. 4a, c, e shows the ensemble-mean RSL trends due to contemporary mass redistribution.

Steric changes

We estimate global-mean and basin-mean steric changes for 1957–2018 from gridded temperature and salinity reconstructions based on in situ observations of temperature and salinity. We use existing gridded estimates[31,32] for the upper 2,000 m. From these observations, we compute steric height anomalies using the TEOS-10 GSW software[61]. We also use gridded steric sea-level change estimates[30]. For each ensemble member, one of these estimates is selected randomly. Before the end of the 1950s, in situ observations are too sparse to derive unbiased steric changes[62]. For the upper-ocean (above 2,000 m) contribution before 1957, we use estimates[15] computed from sea-surface temperature anomaly observations and estimates of ocean heat anomaly pathways from an ocean reanalysis. We also use the deep-ocean (below 2,000 m) steric expansion[15] for the full 1900–2018 period. These estimates come with an uncertainty, which is used to perturb each ensemble member. In the Argo float data, a salinity drift has been detected since 2015[63], which causes an underestimation of global steric sea level. We correct for this drift by removing the estimated global-mean halosteric sea-level changes from each gridded estimate. Extended Data Fig. 7 depicts the time series of the individual steric products and the resulting estimates used in this paper.

Sea-level observations

We use annual-mean tide-gauge observations from the revised local reference (RLR) dataset from the Permanent Service for Mean Sea Level[64,65], as well as an extended tide-gauge dataset[66], which has been updated until 2018. We remove observations that have been flagged for quality issues. Some stations show apparent data problems, such as spikes, jumps, drifts and large trends. These problems are typically caused by earthquakes, local subsidence, levelling issues and instrument problems. Owing to the multitude of the data problems, such stations cannot be automatically flagged and excluded, on the basis of pre-set criteria, and we manually remove these regions from the analysis. We ultimately use 559 individual tide-gauge records in our reconstruction. From each sea-level record, we remove the self-consistent equilibrium nodal cycle[67] and the effects of local wind and sea-level pressure changes. To this end, we use wind and sea-level pressure fields from the ERA-20c reanalysis[68] over 1900–1979 and ERA5 reanalysis[69] from 1979–2018, and use a simple linear regression model to remove the wind and pressure effects[70]. Some locations, such as Aberdeen, Sydney and Singapore, have multiple tide-gauge records with different observational periods. We merge stations that are within 20 km of each other and have an overlap of at least 5 station years into regions. Henceforth, we refer to regions to denote any location that has a single or multiple merged tide-gauge observations. We only consider regions with at least 20 years of data. We link each region to a single ocean basin. All regions and the associated basins are shown in Extended Data Fig. 3.

VLM

Tide-gauge observations are affected by VLM[71], and correcting these records for VLM has resulted in more coherent sea-level trends across different tide gauges[20,72]. We use VLM observations from permanent GNSS stations and from the difference between satellite-altimetry and tide-gauge observations[71,73]. The RSL patterns associated with GIA and GRD are partially caused by solid-Earth deformation, which is observed as VLM. To avoid double-counting, we subtract the modelled solid-Earth deformation that results from GIA (R_{GIA}) and contemporary GRD (R_{GRD}) from the observed VLM time series (R_{obs}), to obtain a time series of residual VLM[9,45]:

$$R_{\mathrm{residual}}(t) = R_{\mathrm{obs}}(t) - R_{\mathrm{GIA}}(t) - R_{\mathrm{GRD}}(t) \tag{1}$$

We compute the linear trend in residual VLM, and we assume that the rate of residual VLM is representative for the full length of the tide-gauge record.

We use the GNSS station database from the University of Nevada, Reno[74]. We select all GNSS stations that are within a 30-km radius of each region, have at least 4 years of daily observations, and for which the standard error of the residual VLM trend does not exceed 1 mm yr^{-1}. We estimate the residual VLM trend using the MIDAS trend estimator[75]. We compute residual VLM for each ensemble member. The uncertainty in the derived trend is caused by the uncertainty in the corrections for GIA and contemporary GRD effects, and by the uncertainty that arises from estimating a linear trend from serially correlated data. The uncertainty due to GIA and contemporary GRD is estimated by computing the residual VLM trend for each individual ensemble member. To account for serial correlation, for each ensemble member we determine the trend uncertainty provided by the MIDAS trend estimator. We then draw a random number from a Gaussian distribution with this trend uncertainty as standard deviation, and perturb the estimated trend with this random number.

To obtain residual VLM trends from the difference between satellite-altimetry and tide-gauge observations, we use the MEaSUREs gridded sea surface height anomalies version 1812 dataset[76]. This dataset has been corrected for calibration issues that caused a sea-level drift over the first years of the altimetry era[34]. The altimetry data covers the period 1993–2018. To obtain local residual VLM, we subtract GIA and contemporary GRD effects from altimetry. We require 15 years of overlap between altimetry and the tide gauge, and select all grid points within a 300-km radius for which the correlation between annual-mean de-trended altimetry and tide-gauge sea level is above 0.5. This value, and the radius of 300 km, are chosen as a compromise between accuracy and the number of locations for which VLM can be estimated[73]. We compute the residual VLM time series for each accepted altimetry grid point, and then compute the mean residual VLM time series by taking the mean of all individual time series, weighted by the correlation with the tide-gauge record. From this time series, we compute the linear trend and standard error by assuming that the serial correlation of the time series can be approximated by a first-order autoregressive process. This computation is performed using the Hector software[77]. For stations for which no single altimetry grid point has a correlation of 0.5 or higher, or for which the standard error is above 1 mm yr^{-1}, no VLM estimate is generated. Similarly to the GNSS approach, we perturb each ensemble member with the trend uncertainty that arises from serial correlation in the time series.

Some VLM observations appear as single outliers compared to nearby other observations, or result in unrealistically high or low sea-level trends. As for the tide-gauge selection procedure, owing to the multitude of possible problems in VLM estimates, no general criteria can be applied to catch these problems. Therefore, we manually remove VLM estimates that show such problems. For regions with multiple GNSS stations, or with both GNSS and altimetry VLM estimates available, we use the average residual VLM trend, weighted by the inverse of the squared standard errors of the individual estimates. We are not able to estimate a VLM trend for all tide-gauge regions. For stations for which no VLM trend is available, we assume no residual VLM and a residual VLM standard error of 1 mm yr^{-1}. This standard error is based on the maximum VLM uncertainty that we accept and on the standard deviation among the residual VLM estimates, 1.5 mm yr^{-1}. In some regions, large sea-level trends are compensated for by large residual VLM trends. As a result, this standard deviation is probably biased high for regions without residual VLM estimates, because regions with a large sea-level trend and no residual VLM estimate are removed during the quality control phase.

Global-mean and basin-mean sea-level reconstruction

Following ref. [9], before merging the individual region estimates into basin-mean curves, we estimate and remove the biases between local sea-level changes in each region and basin-mean sea-level changes that result from GIA, contemporary GRD effects and residual VLM. This correction results in an estimate of basin-mean sea level (η_{basin}), given observed regional sea level (η_{region}), the difference between regional sea-level changes that result from GIA ($\eta_{\mathrm{GIA,region}}$) and GRD ($\eta_{\mathrm{GRD,region}}$), and the associated basin-mean sea-level changes, as well as residual VLM:

$$\begin{aligned}\eta_{\mathrm{basin}}(t) = {} & \eta_{\mathrm{region}}(t) + [\eta_{\mathrm{GIA,basin}}(t) - \eta_{\mathrm{GIA,region}}(t)] \\ & + [\eta_{\mathrm{GRD,basin}}(t) - \eta_{\mathrm{GRD,region}}(t)] + R_{\mathrm{residual}}(t)\end{aligned} \tag{2}$$

Local sea-level variability may not be representative for the basin as a whole. To assess the uncertainty due to this non-representativeness, we perturb each ensemble member of the sea-level observations $\eta_{\mathrm{region}}(t)$ from each individual region with a realization of first-order autoregressive (AR1) noise. The AR1 noise parameters are computed from the standard deviation and the first-order serial correlation of the regional sea-level observations. After computing all basin sea-level estimates from each individual region, we merge all the individual regions into a single basin estimate using the virtual-station method[9,20,78], in which the two nearest regions are merged into a new virtual station halfway between the merged stations. Tide-gauge observations are not tied to a common vertical datum system. To account for different datum systems during the averaging process, we remove the common mean between two series estimated over their overlapping period. This procedure is repeated until one virtual station is left. The sea-level change estimate from the final virtual station is used as the basin-mean estimate. We obtain the final GMSL estimate by averaging the basin-mean estimates, weighted by the relative surface area of each basin.

The resulting GMSL estimate shows a linear trend and multidecadal variability pattern that agree with other recent reconstructions[4,5,20]. These recent reconstructions all show lower twentieth-century rates than do earlier assessments[79,80], as shown in Extended Data Fig. 2.

The global-mean and basin-mean altimetry curves are computed using the same gridded altimetry product as used for the VLM computations. To obtain basin-mean and global-mean RSL, we add the modelled deformation of the seafloor due to GIA and contemporary GRD effects to the altimetry curves[81].

The linear trends and accompanying uncertainty estimates in all basin-mean and global-mean quantities discussed here are computed from the linear trends in each ensemble member. Because the unique

GIA model used in each ensemble member has an associated likelihood, we use the likelihood from the GIA model as the weight for the ensemble member when computing the mean and confidence intervals in all components. Because not all terms follow a Gaussian distribution, the confidence intervals are not assumed to be symmetric, and we directly compute the confidence intervals from the 5th and 95th percentile of the weighted ensemble. We account for the uncertainties due to serial correlation in the time series by adding the estimated trend uncertainty to the ensemble spread in quadrature. We assume that the spectrum of all time series can be approximated by a generalized Gauss–Markov spectrum[82]. We compute the noise parameters and the resulting trend uncertainty using the Hector software[77].

Data availability

The resulting global and basin-scale reconstructions, the time series of global and basin sea-level changes and its contributors, grids with local sea-level and solid-Earth deformation due to contemporary GRD effects, and the individual ensemble members are available at https://doi.org/10.5281/zenodo.3862995.

Code availability

The codes to compute the ensemble of observed sea-level changes and contributing processes, and the post-processing routines to compute statistics and to generate the figures are available at https://github.com/thomasfrederikse/sealevelbudget_20c.

38. Thompson, P. R. & Merrifield, M. A. A unique asymmetry in the pattern of recent sea level change. *Geophys. Res. Lett*. **41**, 7675–7683 (2014).
39. Tamisiea, M. E. Ongoing glacial isostatic contributions to observations of sea level change. *Geophys. J. Int*. **186**, 1036–1044 (2011).
40. Melini, D. & Spada, G. Some remarks on glacial isostatic adjustment modelling uncertainties. *Geophys. J. Int*. **218**, 401–413 (2019).
41. Caron, L. et al. GIA model statistics for GRACE hydrology, cryosphere, and ocean science. *Geophys. Res. Lett*. **45**, 2203–2212 (2018).
42. Peltier, W. R., Argus, D. F. & Drummond, R. Comment on "An assessment of the ICE-6G_C (VM5a) glacial isostatic adjustment model" by Purcell et al. *J. Geophys. Res. Solid Earth* **123**, 2019–2028 (2018).
43. Wiese, D. N., Landerer, F. W. & Watkins, M. M. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Wat. Resour. Res*. **52**, 7490–7502 (2016).
44. Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W. & Luthcke, S. B. Replacing GRACE/GRACE-FO C_{30} with satellite laser ranging: impacts on Antarctic Ice Sheet mass change. *Geophys. Res. Lett*. **47**, e2019GL085488 (2020).
45. Frederikse, T., Landerer, F. W. & Caron, L. The imprints of contemporary mass redistribution on local sea level and vertical land motion observations. *Solid Earth* **10**, 1971–1987 (2019).
46. Pfeffer, W. T. et al. The Randolph Glacier Inventory: a globally complete inventory of glaciers. *J. Glaciol*. **60**, 537–552 (2014).
47. Marzeion, B., Jarosch, A. H. & Hofer, M. Past and future sea-level change from the surface mass balance of glaciers. *Cryosphere* **6**, 1295–1322 (2012).
48. Gardner, A. S. et al. A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. *Science* **340**, 852–857 (2013).
49. Cook, A. J., Fox, A. J., Vaughan, D. G. & Ferrigno, J. G. Retreating glacier fronts on the Antarctic peninsula over the past half-century. *Science* **308**, 541–544 (2005).
50. Smith, J. A. et al. Sub-ice-shelf sediments record history of twentieth-century retreat of Pine Island Glacier. *Nature* **541**, 77–80 (2017).
51. Lehner, B. et al. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ*. **9**, 494–502 (2011).
52. Lettenmaier, D. P. & Milly, P. C. D. Land waters and sea level. *Nat. Geosci*. **2**, 452–454 (2009).
53. Wada, Y. et al. Past and future contribution of global groundwater depletion to sea-level rise. *Geophys. Res. Lett*. **39**, L09402 (2012).
54. Wada, Y. et al. Recent changes in land water storage and its contribution to sea level variations. *Surv. Geophys*. **38**, 131–152 (2017).
55. Tamisiea, M. E. et al. Impact of self-attraction and loading on the annual cycle in sea level. *J. Geophys. Res*. **115**, C07004 (2010).
56. Adhikari, S., Ivins, E. R., Frederikse, T., Landerer, F. W. & Caron, L. Sea-level fingerprints emergent from GRACE mission data. *Earth Syst. Sci. Data* **11**, 629–646 (2019).
57. Schaeffer, N. Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations. *Geochem. Geophys. Geosyst*. **14**, 751–758 (2013).
58. Milne, G. A. & Mitrovica, J. X. Postglacial sea-level change on a rotating Earth. *Geophys. J. Int*. **133**, 1–19 (1998).
59. Wang, H. et al. Load Love numbers and Green's functions for elastic Earth models PREM, iasp91, ak135, and modified models with refined crustal structure from Crust 2.0. *Comput. Geosci*. **49**, 190–199 (2012).
60. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter*. **25**, 297–356 (1981).
61. McDougall, T. J. & Barker, P. M. *Getting Started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox* (SCOR/IAPSO WG127, 2011).
62. Cheng, L. et al. Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv*. **3**, e1601545 (2017).
63. Roemmich, D. et al. On the future of Argo: a global, full-depth, multi-disciplinary array. *Front. Mar. Sci*. **6**, 439 (2019).
64. Holgate, S. J. et al. New data systems and products at the permanent service for mean sea level. *J. Coast. Res*. **29**, 493–504 (2013).
65. Permanent Service for Mean Sea Level (PSMSL). *Tide Gauge Data* (retrieved 29 April 2019); http://www.psmsl.org/data/obtaining/.
66. Hogarth, P. Preliminary analysis of acceleration of sea level rise through the twentieth century using extended tide gauge data sets (August 2014). *J. Geophys. Res. Oceans* **119**, 7645–7659 (2014).
67. Woodworth, P. L. A note on the nodal tide in sea level records. *J. Coast. Res*. **280**, 316–323 (2012).
68. Poli, P. et al. ERA-20C: an atmospheric reanalysis of the twentieth century. *J. Clim*. **29**, 4083–4097 (2016).
69. Copernicus Climate Change Service (C3S). *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate* (2019); https://doi.org/10.24381/cds.f17050d7.
70. Frederikse, T. & Gerkema, T. Multi-decadal variability in seasonal mean sea level along the North Sea coast. *Ocean Sci*. **14**, 1491–1501 (2018).
71. Wöppelmann, G. & Marcos, M. Vertical land motion as a key to understanding sea level change and variability. *Rev. Geophys*. **54**, 64–92 (2016).
72. Wöppelmann, G. et al. Evidence for a differential sea level rise between hemispheres over the twentieth century. *Geophys. Res. Lett*. **41**, 1639–1643 (2014).
73. Kleinherenbrink, M., Riva, R. & Frederikse, T. A comparison of methods to estimate vertical land motion trends from GNSS and altimetry at tide gauge stations. *Ocean Sci*. **14**, 187–204 (2018).
74. Blewitt, G., Hammond, W. & Kreemer, C. Harnessing the GPS data explosion for interdisciplinary science. *Eos* **99**, https://doi.org/10.1029/2018EO104623 (2018).
75. Blewitt, G., Kreemer, C., Hammond, W. C. & Gazeaux, J. MIDAS robust trend estimator for accurate GPS station velocities without step detection. *J. Geophys. Res. Solid Earth* **121**, 2054–2068 (2016).
76. Zlotnicki, V., Qu, Z. & Willis, J. *MEaSUREs Gridded Sea Surface Height Anomalies Version 1812* (PODAAC, 2019); https://doi.org/10.5067/SLREF-CDRV2.
77. Bos, M. S., Fernandes, R. M. S., Williams, S. D. P. & Bastos, L. Fast error analysis of continuous GNSS observations with missing data. *J. Geod*. **87**, 351–360 (2013).
78. Jevrejeva, S., Moore, J., Grinsted, A., Matthews, A. & Spada, G. Trends and acceleration in global and regional sea levels since 1807. *Global Planet. Change* **113**, 11–22 (2014).
79. Church, J. A. & White, N. J. Sea-level rise from the late 19th to the early 21st century. *Surv. Geophys*. **32**, 585–602 (2011).
80. Church, J. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. et al.) Ch. 13, 1137–1216 (Cambridge Univ. Press, 2013).
81. Frederikse, T., Riva, R. E. M. & King, M. A. Ocean bottom deformation due to present-day mass redistribution and its impact on sea level observations. *Geophys. Res. Lett*. **44**, 12306–12314 (2017).
82. Langbein, J. Noise in two-color electronic distance meter measurements revisited. *J. Geophys. Res. Solid Earth* **109**, B04406 (2004).

Acknowledgements All figures were made using Generic Mapping Tools (GMT). Parts of this research (T.F., F.L., S.A., L. Caron) were conducted at the Jet Propulsion Laboratory, which is operated for NASA under contract with the California Institute of Technology. S.D. acknowledges the University of Siegen for funding a research stay at JPL. L. Cheng is supported by National Key R&D Program of China (2017YFA0603200).

Author contributions T.F. and F.L. conceived and designed the study. L. Caron and S.A. provided the GIA data and provided guidance on the solid-Earth processes. D.P. provided glacier datasets and helped interpret the underlying uncertainties. V.W.H. provided the TWS reconstruction. P.H. prepared the tide-gauge dataset. L.Z. and L. Cheng helped analyse the steric datasets. Y.-H.W. created the reservoir databases. S.D. provided guidance on the sea-level reconstruction approach. T.F. performed the analysis and wrote the manuscript. All authors contributed to the discussion and helped write the manuscript.

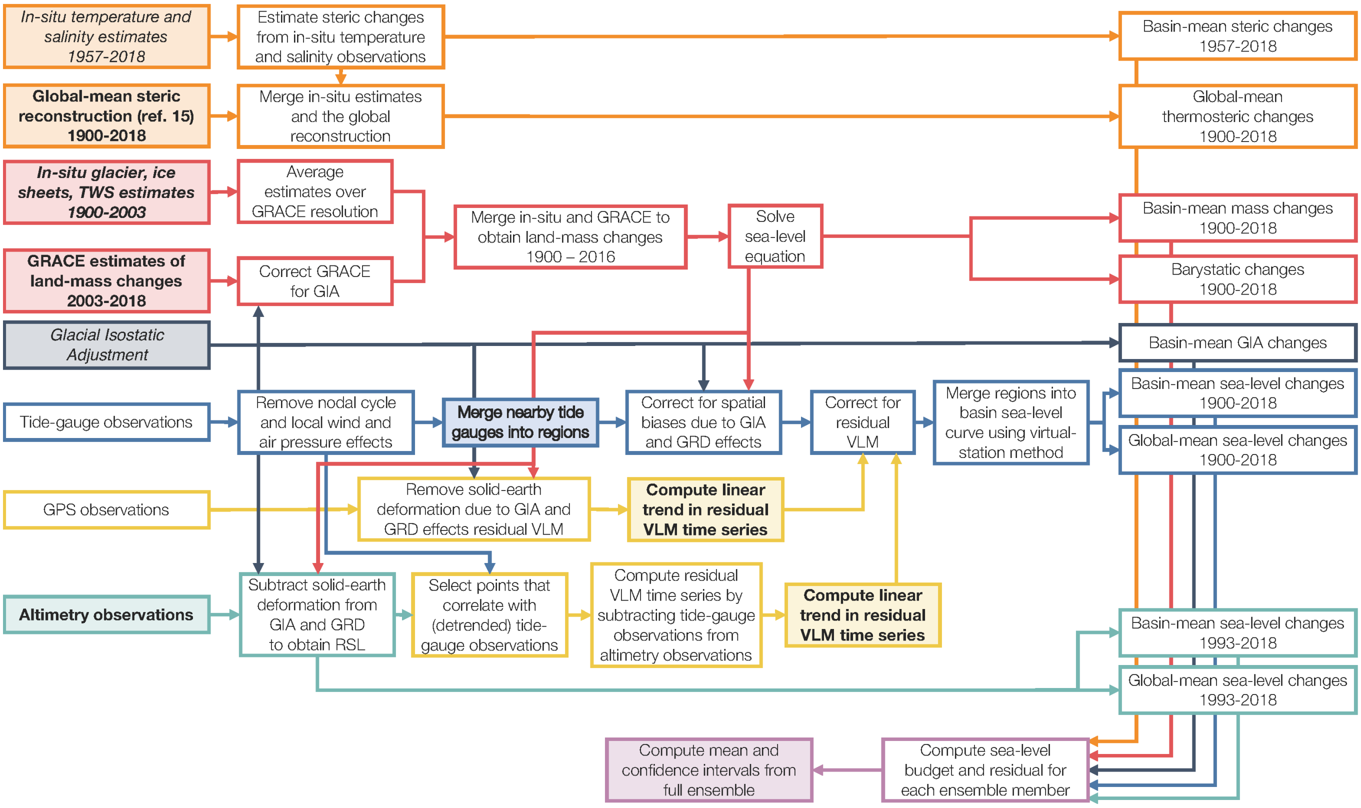
Competing interests The authors declare no competing interests.

Additional information

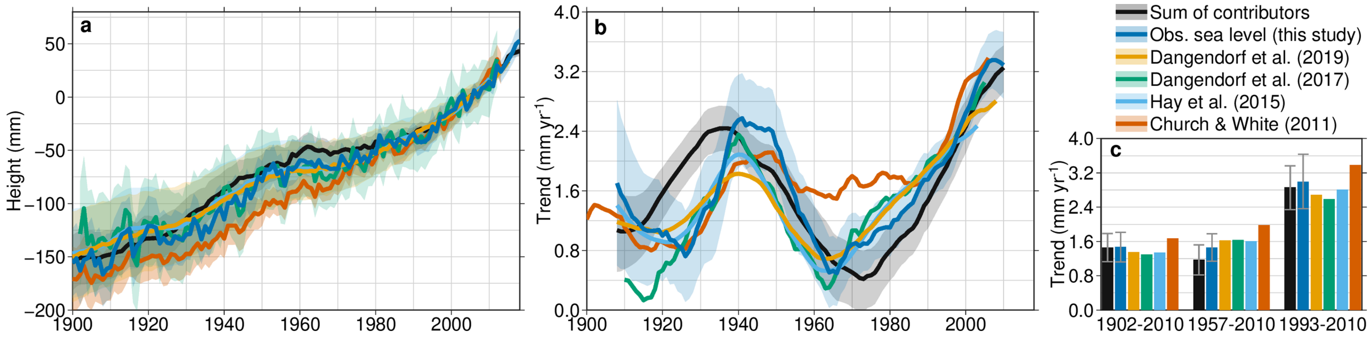
Correspondence and requests for materials should be addressed to T.F.

Peer review information *Nature* thanks Benoît Meyssignac and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

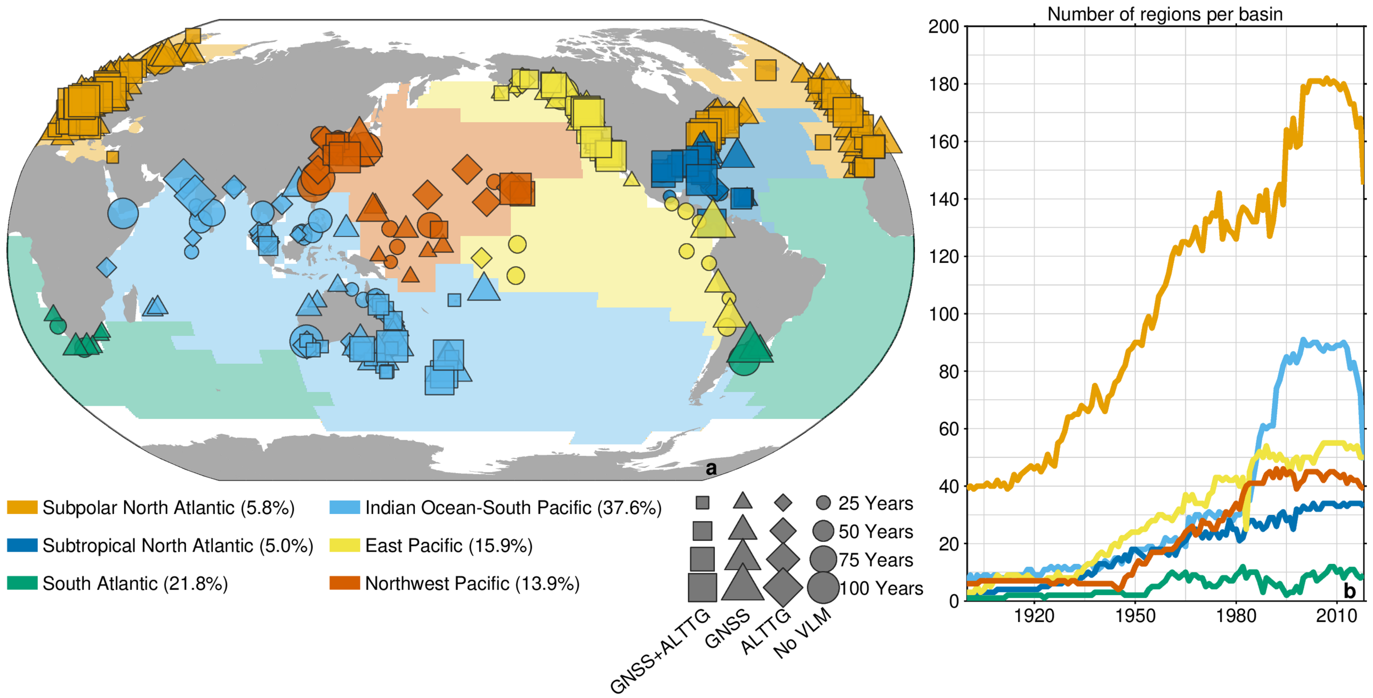
Reprints and permissions information is available at http://www.nature.com/reprints.



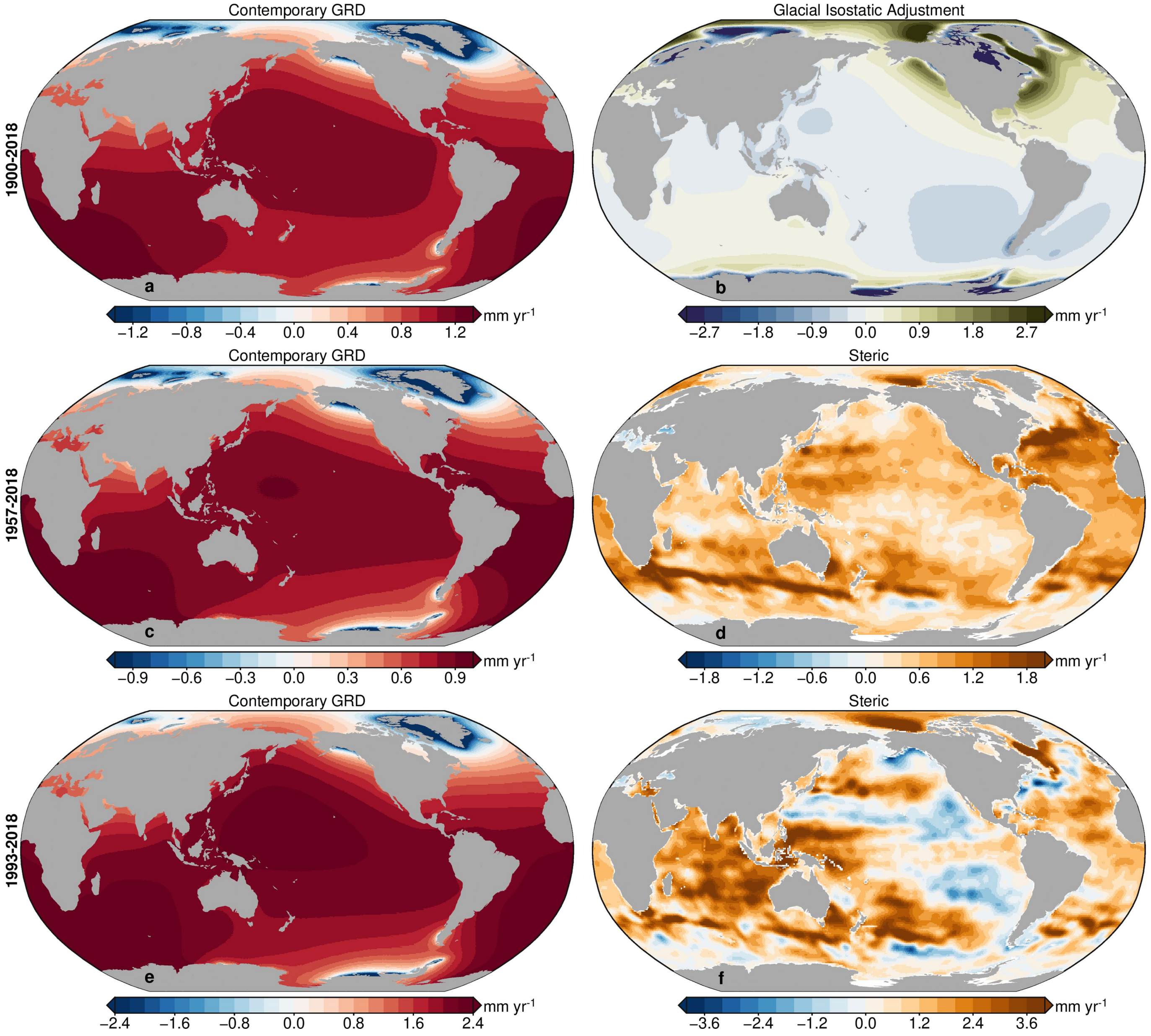
Extended Data Fig. 1 | Schematic overview of the computation of reconstructed sea level and the contributors. Steps with a shaded background involve steps where each ensemble member is perturbed: for steps shown in bold, the estimate is drawn from a probability density function; for steps in italic, the estimate is randomly chosen from a pool of estimates. All the steps are repeated for each of the 5,000 ensemble members, until the final step, where all ensemble members are combined to estimate the mean and confidence interval of the global-mean and basin-mean sea-level budget and its components. Steps and arrows in orange refer to estimates of steric sea level, red refers to estimates of ocean mass, dark blue refers to GIA, light blue refers to tide-gauge observations, yellow refers to VLM, turquoise refers to satellite altimetry, and purple refers to the budget analysis. The global-mean steric reconstruction for 1900–2018 is from ref. [15].



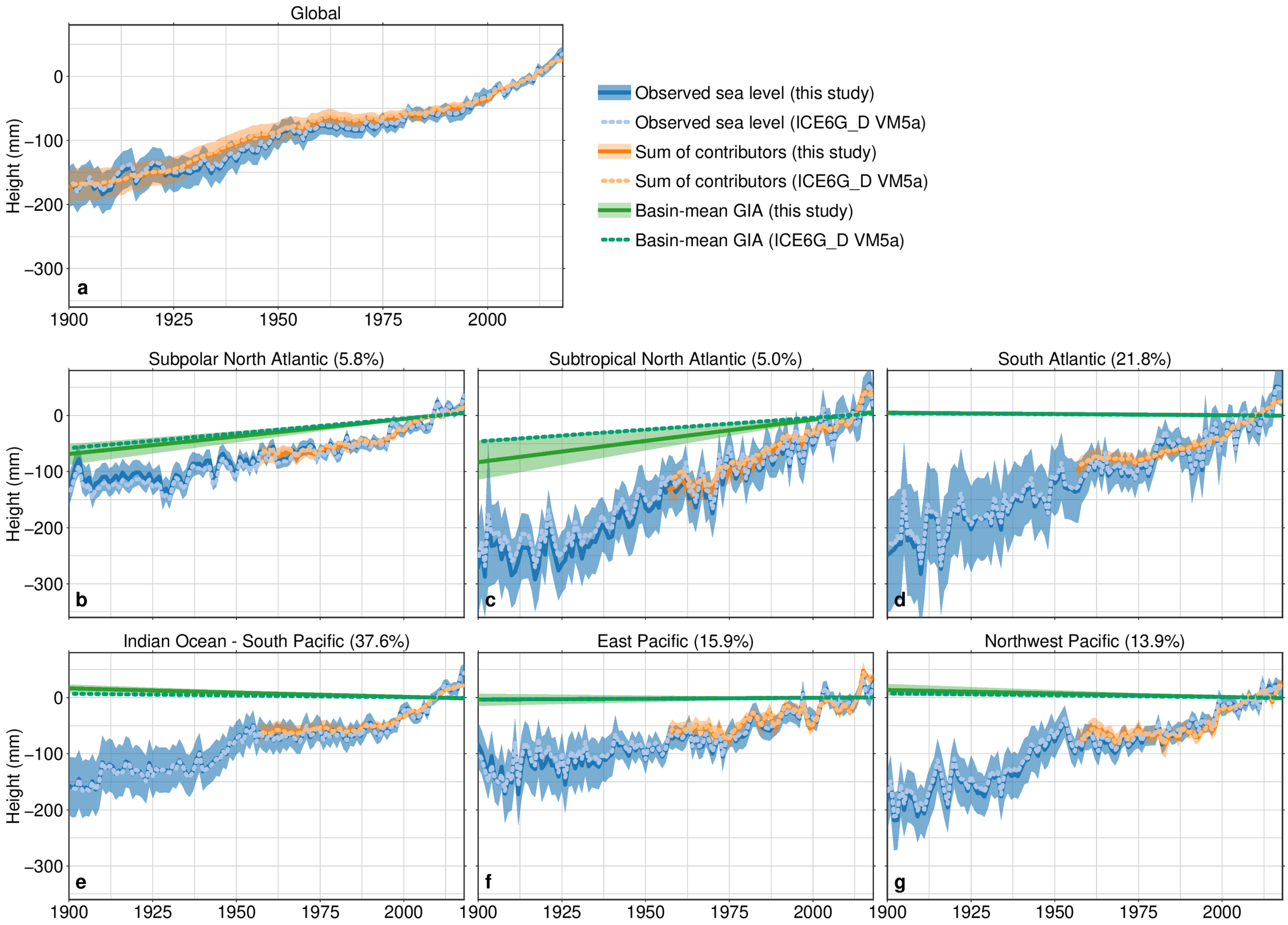
Extended Data Fig. 2 | Comparison with other recent sea-level reconstructions. All panels show observed GMSL and the sum of contributors from this study and other recent GMSL reconstructions[4,5,8,20]. **a**, Annual time series and their 90% confidence intervals. **b**, 30-year-average rates of the GMSL reconstructions. **c**, Linear trends over the time intervals indicated. The shaded regions denote the 90% confidence interval. The values are relative to the 2002–2018 mean.



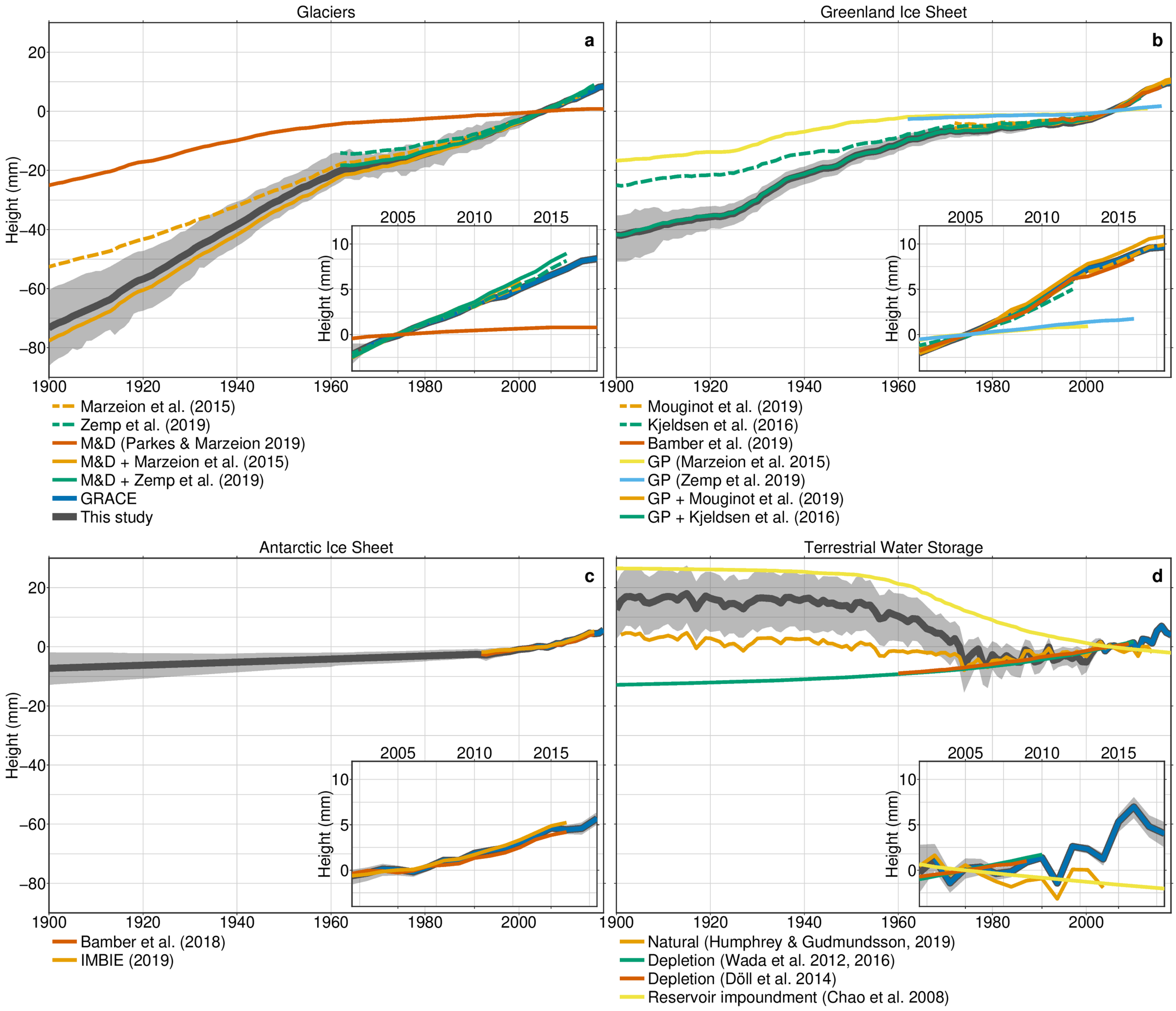
Extended Data Fig. 3 | Map of the regions and basins. a, The ocean basins (shading) and the regions (symbols) that belong to each basin. The shape of the region symbols denotes how the station is corrected for VLM; the size denotes the number of years for which the region provides data. The percentages in the legend show the relative size of the basin as a fraction of the sum of all basins. Each region consists of one tide-gauge station or multiple stations within a 10-km radius. **b**, The number of regions that provide data in a given year for each basin.



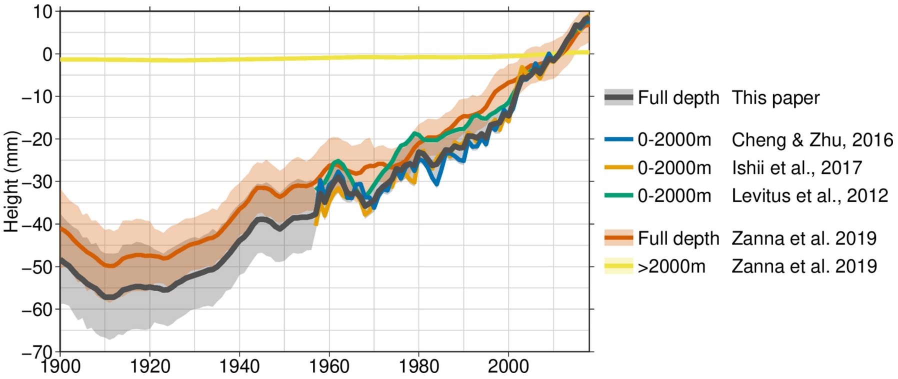
Extended Data Fig. 4 | Linear trends in regional RSL due to ocean-mass changes, GIA and steric changes over three periods. a, **c**, **e**, Local RSL trends due to contemporary GRD effects, for 1900–2018 (**a**), 1957–2018 (**c**) and 1993–2018 (**e**). **b**, RSL changes due to GIA. **d**, **f**, Local steric sea-level changes over 1957–2018 (**d**) and 1993–2018 (**f**). All trends show the ensemble-mean values. The colour scale varies between panels.



Extended Data Fig. 5 | Comparison of two GIA models. Each panel shows observed sea level, the sum of contributors and the basin-mean sea-level trend due to GIA, using the model used in this study[41] (solid lines) and using the ICE6G D (VM5a) model[42]. Shaded areas indicate 90% confidence intervals. **a**, GMSL. **b**–**g**, Basin-mean sea level.



Extended Data Fig. 6 | Central values of individual estimates and our composite final estimate of each barystatic contributor. a, The glacier estimates; data from this study and refs. [16,18,21]. M&D, missing and disappeared glaciers (as determined by ref. [16]). **b**, The Greenland Ice Sheet estimates; data from refs. [14,18,21,24,25]. GP, contribution from Greenland peripheral glaciers. **c**, The Antarctic Ice Sheet estimates; data from refs. [23,24]. **d**, The TWS estimates; data from refs. [17,26–28,53]. All estimates are shown relative to the average height over 2003–2005. The inset in each panel shows all the estimates over 2002–2018. Shaded areas indicate 90% confidence intervals.



Extended Data Fig. 7 | Individual estimates of global-mean steric sea-level changes. The coloured time series show global-mean steric sea-level changes of each individual estimate[15,30–32] and the averaged estimate used here (black). The shaded areas indicate 90% confidence intervals.

Extended Data Table 1 | Trends in observed basin-mean sea level and its contributors

Subpolar North Atlantic	1900-2018		1957-2018		1993-2018	
Glaciers	0.42	[0.32 0.52]	0.31	[0.21 0.41]	0.36	[0.28 0.47]
Greenland Ice Sheet	-0.16	[-0.19 -0.13]	-0.11	[-0.14 -0.08]	-0.25	[-0.27 -0.22]
Antarctic Ice Sheet	0.08	[-0.01 0.17]	0.13	[0.04 0.22]	0.33	[0.21 0.45]
Terrestrial Water Storage	-0.14	[-0.23 -0.06]	-0.08	[-0.20 0.03]	0.16	[0.04 0.28]
Barystatic	0.19	[0.05 0.34]	0.25	[0.09 0.41]	0.61	[0.40 0.82]
Glacial Isostatic Adjustment	0.62	[0.47 0.80]	0.62	[0.47 0.80]	0.62	[0.47 0.80]
Steric	-		0.62	[0.40 0.86]	1.18	[1.00 1.37]
Sum of Contributors	-		1.50	[1.10 1.92]	2.42	[2.00 2.82]
Observed sea level	1.08	[0.79 1.38]	1.52	[1.23 1.83]	2.69	[2.18 3.18]
Altimetry	-		-		2.17	[1.66 2.66]

Indian Ocean-South Pacific	1900-2018		1957-2018		1993-2018	
Glaciers	0.73	[0.55 0.90]	0.56	[0.35 0.72]	0.73	[0.56 0.89]
Greenland Ice Sheet	0.48	[0.39 0.58]	0.33	[0.24 0.42]	0.72	[0.63 0.80]
Antarctic Ice Sheet	0.05	[-0.02 0.12]	0.09	[0.02 0.16]	0.22	[0.13 0.32]
Terrestrial Water Storage	-0.24	[-0.37 -0.11]	-0.17	[-0.34 -0.01]	0.30	[0.11 0.48]
Barystatic	1.03	[0.73 1.34]	0.79	[0.47 1.11]	1.97	[1.61 2.32]
Glacial Isostatic Adjustment	-0.15	[-0.21 -0.08]	-0.15	[-0.21 -0.08]	-0.15	[-0.21 -0.08]
Steric	-		0.64	[0.41 0.87]	1.50	[1.26 1.76]
Sum of Contributors	-		1.29	[0.68 1.91]	3.32	[2.68 3.94]
Observed sea level	1.33	[0.80 1.86]	1.51	[1.03 2.00]	3.93	[3.32 4.55]
Altimetry	-		-		3.65	[3.23 4.08]

Subtropical North Atlantic	1900-2018		1957-2018		1993-2018	
Glaciers	0.68	[0.50 0.85]	0.50	[0.30 0.65]	0.62	[0.46 0.77]
Greenland Ice Sheet	0.23	[0.18 0.27]	0.15	[0.11 0.20]	0.34	[0.29 0.38]
Antarctic Ice Sheet	0.10	[-0.01 0.20]	0.16	[0.06 0.26]	0.40	[0.26 0.52]
Terrestrial Water Storage	-0.13	[-0.23 -0.03]	-0.05	[-0.18 0.09]	0.28	[0.13 0.43]
Barystatic	0.87	[0.62 1.12]	0.77	[0.50 1.03]	1.63	[1.33 1.92]
Glacial Isostatic Adjustment	0.76	[0.40 1.04]	0.76	[0.40 1.04]	0.76	[0.40 1.04]
Steric	-		1.29	[1.02 1.58]	1.08	[0.60 1.50]
Sum of Contributors	-		2.81	[2.29 3.35]	3.48	[2.72 4.19]
Observed sea level	2.49	[1.89 3.06]	2.76	[2.05 3.42]	3.98	[2.75 5.20]
Altimetry	-		-		4.04	[2.77 5.24]

East Pacific	1900-2018		1957-2018		1993-2018	
Glaciers	0.66	[0.47 0.85]	0.48	[0.25 0.64]	0.62	[0.44 0.76]
Greenland Ice Sheet	0.48	[0.39 0.58]	0.33	[0.24 0.42]	0.72	[0.63 0.81]
Antarctic Ice Sheet	0.09	[-0.01 0.19]	0.15	[0.05 0.24]	0.37	[0.24 0.49]
Terrestrial Water Storage	-0.22	[-0.34 -0.09]	-0.15	[-0.32 0.02]	0.32	[0.13 0.49]
Barystatic	1.02	[0.70 1.32]	0.81	[0.46 1.13]	2.02	[1.66 2.37]
Glacial Isostatic Adjustment	0.03	[-0.06 0.13]	0.03	[-0.06 0.13]	0.03	[-0.06 0.13]
Steric	-		0.47	[0.21 0.74]	0.37	[-0.01 0.77]
Sum of Contributors	-		1.32	[0.86 1.74]	2.43	[1.90 2.95]
Observed sea level	1.20	[0.76 1.62]	1.64	[1.26 2.03]	1.82	[1.10 2.56]
Altimetry	-		-		2.35	[0.70 4.06]

South Atlantic	1900-2018		1957-2018		1993-2018	
Glaciers	0.76	[0.56 0.95]	0.56	[0.32 0.73]	0.72	[0.54 0.88]
Greenland Ice Sheet	0.50	[0.41 0.60]	0.34	[0.24 0.43]	0.74	[0.65 0.83]
Antarctic Ice Sheet	0.09	[-0.01 0.19]	0.16	[0.06 0.25]	0.38	[0.26 0.50]
Terrestrial Water Storage	-0.20	[-0.35 -0.05]	-0.12	[-0.30 0.07]	0.38	[0.18 0.58]
Barystatic	1.15	[0.82 1.48]	0.93	[0.58 1.26]	2.23	[1.87 2.61]
Glacial Isostatic Adjustment	-0.04	[-0.08 -0.02]	-0.04	[-0.08 -0.02]	-0.04	[-0.08 -0.02]
Steric	-		0.88	[0.73 1.03]	1.29	[0.98 1.66]
Sum of Contributors	-		1.78	[1.23 2.33]	3.48	[2.79 4.15]
Observed sea level	2.07	[1.36 2.77]	2.17	[1.62 2.73]	3.89	[2.44 5.33]
Altimetry	-		-		3.45	[3.04 3.86]

Northwest Pacific	1900-2018		1957-2018		1993-2018	
Glaciers	0.69	[0.50 0.88]	0.50	[0.28 0.67]	0.65	[0.47 0.80]
Greenland Ice Sheet	0.53	[0.42 0.64]	0.35	[0.26 0.45]	0.78	[0.68 0.88]
Antarctic Ice Sheet	0.10	[-0.01 0.20]	0.16	[0.06 0.26]	0.40	[0.26 0.53]
Terrestrial Water Storage	-0.22	[-0.36 -0.09]	-0.16	[-0.34 0.02]	0.34	[0.15 0.52]
Barystatic	1.09	[0.76 1.41]	0.86	[0.51 1.21]	2.17	[1.78 2.53]
Glacial Isostatic Adjustment	-0.12	[-0.22 -0.03]	-0.12	[-0.22 -0.03]	-0.12	[-0.22 -0.03]
Steric	-		0.71	[0.45 0.98]	1.20	[0.79 1.58]
Sum of Contributors	-		1.44	[0.80 2.07]	3.23	[2.52 3.93]
Observed sea level	1.68	[1.27 2.09]	1.80	[1.42 2.18]	2.77	[2.11 3.39]
Altimetry	-		-		3.53	[2.64 4.45]

The trends are given in millimetres per year, over 1900–2018, 1957–2018 and 1993–2018. The numbers in brackets indicate the 90% confidence intervals.

Zoonotic host diversity increases in human-dominated ecosystems

https://doi.org/10.1038/s41586-020-2562-8

Received: 28 January 2019

Accepted: 1 July 2020

Rory Gibb[1,5], David W. Redding[1,5] ✉, Kai Qing Chin[1], Christl A. Donnelly[2,3], Tim M. Blackburn[1,4], Tim Newbold[1] & Kate E. Jones[1,4] ✉

Land use change—for example, the conversion of natural habitats to agricultural or urban ecosystems—is widely recognized to influence the risk and emergence of zoonotic disease in humans[1,2]. However, whether such changes in risk are underpinned by predictable ecological changes remains unclear. It has been suggested that habitat disturbance might cause predictable changes in the local diversity and taxonomic composition of potential reservoir hosts, owing to systematic, trait-mediated differences in species resilience to human pressures[3,4]. Here we analyse 6,801 ecological assemblages and 376 host species worldwide, controlling for research effort, and show that land use has global and systematic effects on local zoonotic host communities. Known wildlife hosts of human-shared pathogens and parasites overall comprise a greater proportion of local species richness (18–72% higher) and total abundance (21–144% higher) in sites under substantial human use (secondary, agricultural and urban ecosystems) compared with nearby undisturbed habitats. The magnitude of this effect varies taxonomically and is strongest for rodent, bat and passerine bird zoonotic host species, which may be one factor that underpins the global importance of these taxa as zoonotic reservoirs. We further show that mammal species that harbour more pathogens overall (either human-shared or non-human-shared) are more likely to occur in human-managed ecosystems, suggesting that these trends may be mediated by ecological or life-history traits that influence both host status and tolerance to human disturbance[5,6]. Our results suggest that global changes in the mode and the intensity of land use are creating expanding hazardous interfaces between people, livestock and wildlife reservoirs of zoonotic disease.

Anthropogenic environmental change affects many dimensions of human health and wellbeing, including the incidence and emergence of zoonotic and vector-borne diseases[1]. Although large-scale research into environmental drivers of disease has mostly focused on climate, there is a growing consensus that land use change—the conversion of natural habitats to agricultural, urban or otherwise anthropogenic ecosystems—is a globally important mediator of infection risk and disease emergence in humans[2,3]. Land use change directly and indirectly drives the loss, turnover and homogenization of biodiversity (including through invasions and rare species losses)[7,8], modifies the structure of the landscape in ways that modulate epidemiological processes (for example, fragmentation[9] and resource provisioning[10]) and can increase contact between humans and wildlife (for example, through agricultural practices and hunting)[1]. These processes interact to influence transmission dynamics in reservoir and vector communities and, ultimately, pathogen spillover risk to humans[11,12], with land use change implicated in driving both endemic (for example, trypanosomiasis[13] and malaria[14]) and epidemic (for example, Nipah[15] and West Nile[16]) zoonoses. However, the complexity of these systems (Extended Data Fig. 1) has made it difficult to identify whether land use has consistent effects on the ecological factors that underpin zoonotic disease risk[2]—a critical knowledge gap given the ongoing trends in global land use change[17].

Although there is broad evidence for regulatory effects of local species diversity on pathogen transmission[18], such effects are not universal: higher disease risk in depauperate assemblages has been observed for some disease systems (for example *Borrelia*[19], West Nile[16] and *Ribeiroia*[6]) but not others. One ecological factor underlying these inconsistencies might be differences in the sensitivity of host species to human pressures[4]. It is often proposed that more effective zoonotic host species might be generally more likely to persist in disturbed ecosystems, because certain trait profiles (for example, 'fast' life histories and higher population densities) correlate with both reservoir status and reduced extirpation risk in several vertebrate taxa[20,21]. Alternatively, any such tendencies might be taxonomically or geographically idiosyncratic: for example, mammals that are more closely phylogenetically related to humans are more likely to be zoonotic reservoirs[22], but might also

---

[1] Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK. [2] Department of Statistics, University of Oxford, Oxford, UK. [3] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK. [4] Institute of Zoology, Zoological Society of London, London, UK. [5] These authors contributed equally: Rory Gibb, David W. Redding. ✉e-mail: d.redding@ucl.ac.uk; kate.e.jones@ucl.ac.uk

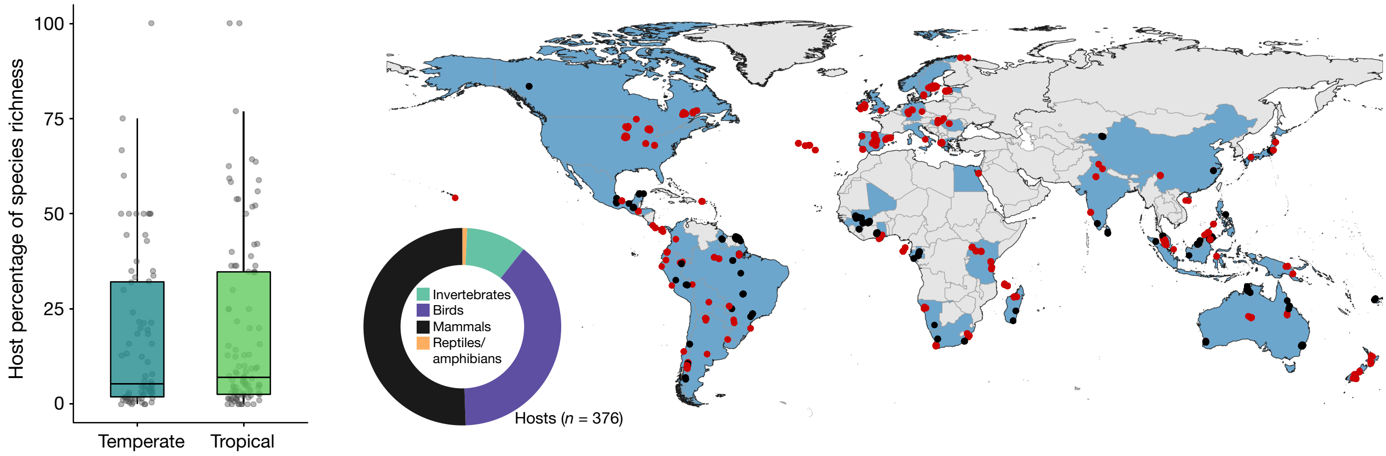


Fig. 1 | Dataset of ecological communities and zoonotic host species. Points on the map show the geographical locations of surveyed assemblages ($n = 6,801$ sites), with mammal survey locations in black and all other sites in red, and countries containing sites shaded in blue. The chart shows the taxonomic distribution of hosts of human-shared pathogens (birds, invertebrates, mammals, reptiles and amphibians; see Methods). Box plots and points show, for each study, host species richness as a percentage of the total per-study sampled richness, split across temperate and tropical biomes ($n = 184$ studies; boxes show median and interquartile range (IQR), whiskers show values within 1.5 × IQR of quartiles). Map generated using Natural Earth (https://naturalearthdata.com).

be highly variable in their sensitivity to anthropogenic disturbance[20]. Responses of reservoir hosts to disturbance have been investigated in certain taxa (for example, primates[23]) and disease systems[13,19], but so far there has been no comprehensive analysis of the effects of land use on zoonotic host diversity and species composition.

Here we use a global dataset of 6,801 ecological assemblages derived from the Projecting Responses of Ecological Diversity in Changing Terrestrial Systems (PREDICTS) biodiversity database[24] to test whether land use has systematic effects on the zoonotic potential of wildlife communities. We identified records of wildlife hosts of known human pathogens and endoparasites (henceforth referred to as 'pathogens') within PREDICTS using a comprehensive host–pathogen associations database, and classified species as zoonotic hosts (henceforth 'hosts') on the basis of evidence of association with at least one human-shared pathogen (see Methods). PREDICTS compiles more than 3.2 million species records from 666 published studies that sampled biodiversity across land use gradients using consistent protocols, enabling a global comparison of local assemblages in primary vegetation (minimally disturbed baseline) to nearby secondary (recovering from past disturbance), managed (cropland, pasture or plantation) and urban sites, of varying use intensities (here, minimal or substantial use)[24]. We identified records of 376 host species in a dataset of 6,801 survey sites from 184 studies across 6 continents, with a taxonomic distribution broadly representative of known zoonotic host diversity (Fig. 1, Supplementary Tables 1, 2; Methods). Host responses to land use were compared with the responses of all other species at the same locations (termed 'non-hosts', approximating the response of background biodiversity; $n = 6,512$ species), using Bayesian mixed-effects models to control for study methods and sampling design (Methods). Pathogen detection is sensitive to research effort, such that some poorly studied species might be misclassified as non-hosts. We account for this uncertainty in our models using a bootstrap approach, in which each iteration transitions a proportion of non-host species to host status, with species-level transition rates determined by both publication effort and taxonomic order (Supplementary Methods 1, Extended Data Fig. 2). All parameter estimates are obtained across each full bootstrap ensemble (Methods).

We first estimated the effects of land use type and intensity on two community metrics: site-level host species richness (number of host species; related to potential pathogen richness) and host total abundance (total number of host individuals; a more epidemiologically relevant metric related to opportunities for transmission)[25]. Both host richness and total abundance either persist or increase in response to land use, against a background of consistent declines in all other (non-host) species in human-dominated habitats (Fig. 2a, b). Together these changes result in hosts comprising an increasing proportion of ecological assemblages in secondary, managed and urban land (Fig. 2c, d, Supplementary Tables 3–5). Notably, land use intensity has clear positive effects on community zoonotic potential both within and between land use types, with the largest increases seen for substantial-use secondary and managed sites (posterior median: +18–21% host proportion richness, +21–26% proportion abundance) and urban sites (+62–72% proportion richness, +136–144% proportion abundance; but with higher uncertainty due to sparser sampling). These results are robust to testing for sensitivity to random study-level variability (Extended Data Fig. 3a), geographical biases in data coverage[24] (Extended Data Fig. 3b) and strictness of host status definition (Extended Data Fig. 4). The latter of these is crucial to understanding disease risk, because species that are capable of being infected by a given pathogen might not contribute substantially to transmission dynamics or zoonotic spillover risk. We therefore repeated the analyses using a stricter reservoir host definition, focusing on mammals as they are the major reservoirs of zoonoses globally. We strictly defined reservoir status as an association with at least one zoonotic agent (an aetiological agent of a specific human disease with a known animal reservoir), and defined association on the basis of detection or isolation of the pathogen, or confirmed reservoir status. In total, 143 host species, 2,026 sites and 63 studies were considered. The overall trends remained consistent, although with notably stronger effects on host proportion of total abundance (+42–52% in secondary and managed land), and weaker effects on host richness that may reflect underlying variability in responses between mammal taxa (Extended Data Fig. 4).

To examine the possibility of such taxonomic variability in host responses to land use, we analysed mean land use effects on species-level occurrence and abundance of zoonotic host (strictly defined) and non-host species, for several mammalian (Carnivora, Cetartiodactyla, Chiroptera, Primates, Rodentia) and avian (Passeriformes, Psittaciformes) orders that are well-sampled in PREDICTS and harbour the majority of known zoonoses (Methods). Within most orders, non-host species tend to decline more strongly in response to land disturbance than do host species, but with substantial between-order variation in the direction and clarity of effects (Fig. 3, Extended Data Fig. 5, Supplementary Table 6). Notably, within passerine birds, bats and

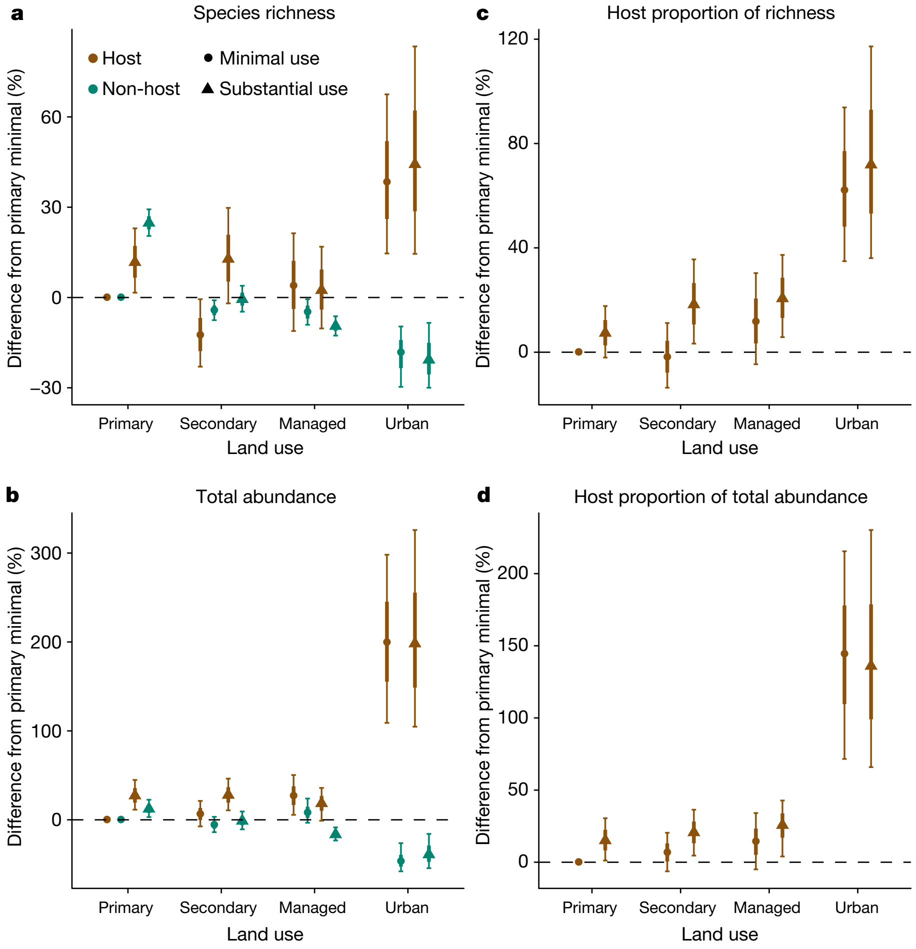


Fig. 2 | Effects of land use on site-level host species richness and total abundance. a–**d**, Models of species richness (**a**) and total abundance (**b**) of host species and of all other (non-host) species, and of hosts as a proportion of total site-level richness (**c**) and abundance (**d**). Points, wide and narrow error bars show modelled percentage difference in diversity metrics (posterior marginal median, 67% and 95% quantile ranges, respectively, across 1,000 bootstrap models) relative to a baseline of primary land under minimal use (dashed line) ($n = 6,801$ sites: primary (1,423 and 1,457 for minimal and substantial use, respectively), secondary (1,044, 629), managed (565, 1,314), urban (136, 233)). All posterior estimates were calculated across an ensemble of 1,000 bootstrapped models, each with a proportion of non-hosts probabilistically transitioned to host status (median 121, range 90–150; Extended Data Fig. 2) to account for variability in species-level research effort (Methods, Supplementary Methods 1). Models also included fixed effects for human population density and random effects for study methods and biome (Methods). Parameter estimates represent average effect sizes across multiple studies with differing survey methods and taxonomic focus, so do not have an absolute numerical interpretation.

rodents, hosts and non-hosts show clear divergent responses to land use, with abundances of host species on average increasing (Passeriformes, +14–96%; Chiroptera, +45%; Rodentia, +52%) while abundances of non-host species decline (Passeriformes, −28% to −43%; Chiroptera, −13%; Rodentia, −53%) in human-dominated sites relative to primary sites (Fig. 3). Although such a tendency has been observed in some disease systems, our results suggest that this is a more general phenomenon in these taxa, which may contribute to numerous documented links between anthropogenic ecosystems and bat-, rodent- and bird-borne emerging infections (for example, corona-, henipa-, arena- and flaviviruses, *Borrelia* and *Leptospira* spp.)[15,16,19]. By contrast, primate and carnivore host responses are not clearly distinguishable from overall species declines in these orders; this is consistent with past studies that show no consistent links between land disturbance and disease in primates[23], and highlights the importance of ecotonal or edge habitats as epidemiological interfaces between humans and primates[14] (although sparser urban sampling means that urban-adapted primates, such as macaques, are likely to be underrepresented).

The differing responses of host and non-host species may be mediated by covariance between traits that influence both host status and human tolerance[26], but could also reflect histories of human–wildlife contact and coevolution of shared pathogens[11]. If the former is the case, we expect that harbouring a higher number of pathogens overall (richness of either zoonotic or non-zoonotic pathogens; a metric often correlated with species traits[27]), would be associated with more positive species responses to land use. We tested this across all mammals in our dataset (owing to more complete pathogen data availability than for other taxa; 546 species, 1,950 sites), here controlling for species-level differences in research effort by analysing residual pathogen richness not explained by publication effort (Methods, Extended Data Fig. 6). We find that pathogen richness is associated with increasing probability of species occurrence in managed sites but not in primary habitat, and that this result is consistent for either human-shared or non-human-shared pathogens (no documented infection of either people or domestic animals; Extended Data Fig. 7, Supplementary Table 7). This suggests that the net increase in zoonotic host diversity in disturbed sites is at least partly trait-mediated; in particular, species traits associated with a faster pace of life are often correlated both with reservoir status and with infection outcomes[5,26] (potentially owing to life-history trade-offs between reproductive rate and immune investment[28]), and with resilience to anthropogenic pressures[20].

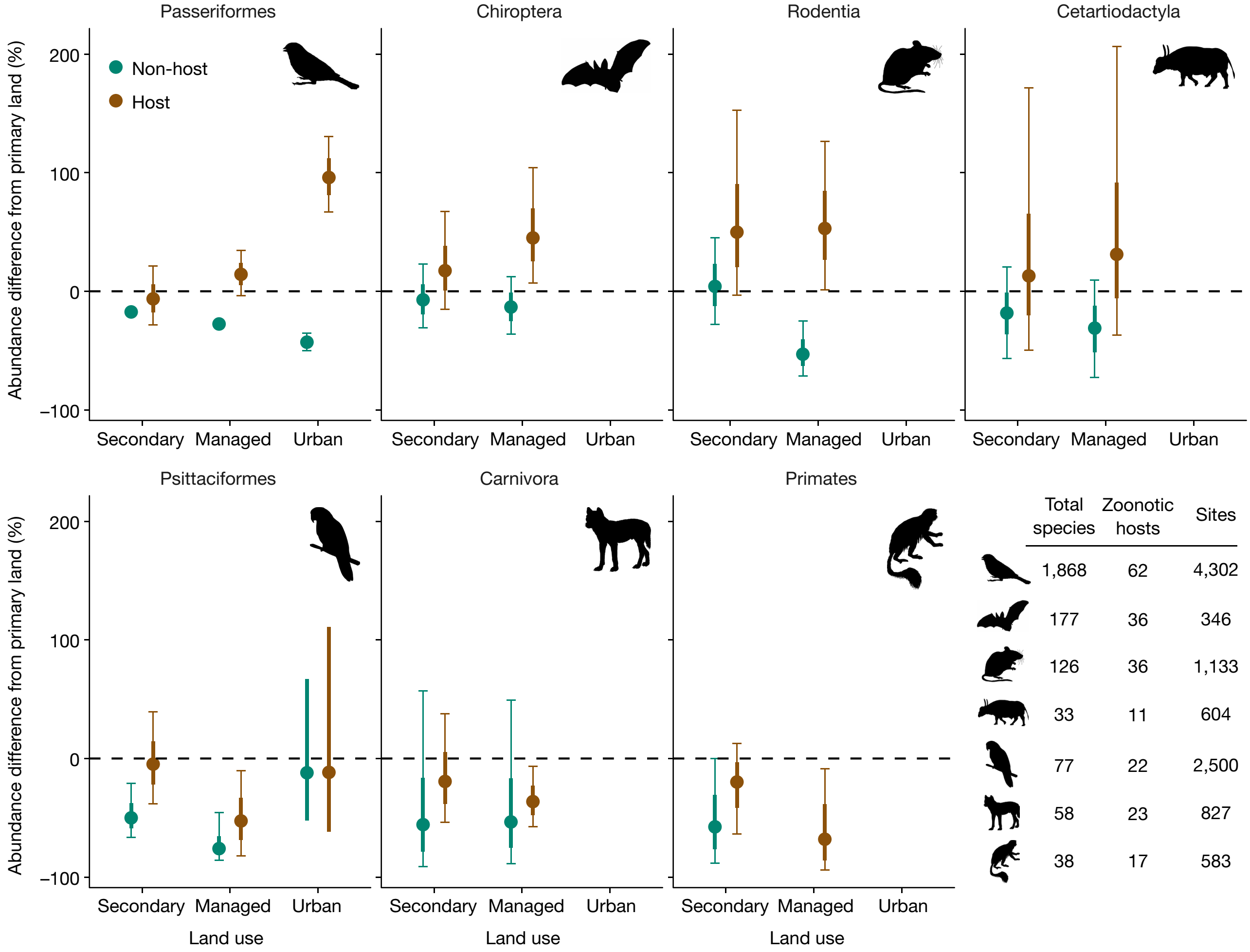


Fig. 3 | Effects of land use on species abundance of mammalian and avian zoonotic hosts and non-hosts. Points, wide and narrow error bars show average difference in species abundance (posterior median, 67% and 95% quantile ranges, respectively, across 500 bootstrap models to account for host status uncertainty) in secondary, managed and urban sites relative to a primary land baseline (dashed line). Differences are estimated across all host and non-host species in each mammalian or avian order. For mammals, zoonotic host status was defined strictly (direct pathogen detection, isolation or confirmed reservoir status), and urban sites were excluded owing to sparse urban sampling (only two studies; in addition, no non-host primates were recorded in managed land, and urban 95% quantile range for Psittaciformes is not shown owing to high uncertainty). Abundance differences were predicted using a hurdle-model-based approach to account for zero-inflation (combining separately fitted occurrence and zero-truncated abundance models; see Extended Data Fig. 5, Methods). The table shows per-order numbers of species in the dataset (between 8% and 35% of the total described species in each order), known zoonotic hosts (before bootstrap) and sampled sites. Silhouettes obtained from PhyloPic (http://phylopic.org/).

A trait-mediated explanation is also supported by our finding that differential host and non-host species responses to land use are most clearly detected when comparing across large clades with a wide diversity of life histories—such as rodents, passerines and, notably, mammals overall (Extended Data Fig. 5). By contrast, clades that are generally longer-lived and larger-bodied (for example, primates and carnivores) show more idiosyncratic or negative responses to landscape disturbance (Fig. 3).

Overall, our results indicate that the homogenizing effects of land use on biodiversity globally[8] have produced systematic changes to local zoonotic host communities, which may be one factor underpinning links between human-disturbed ecosystems and the emergence of disease. By leveraging site-level survey data, our analyses reflect community changes at the epidemiologically relevant local-landscape scale[21], negating the need to ignore community interactions or generalize ecological processes to coarser spatial scales (a typical limitation of global studies that can confound or mask biodiversity–disease relationships[29]). Our results reflect potential zoonotic hazard, because proximity to reservoir hosts is not sufficient for spillover[30] and emergent disease risk will depend on contextual factors (for example, pathogen prevalence, intermediate host and vector populations, landscape structure, socioeconomics) that may synergistically or antagonistically affect transmission dynamics and exposure rates[11]. Nonetheless, land use also predictably affects other factors that can amplify within-species and cross-species transmission[31] (such as resource provisioning[10] and vector diversity[32]), and increases the potential for human–wildlife contact[12]: for example, human populations are consistently higher at disturbed sites in our dataset (Extended Data Fig. 8). The global expansion of agricultural and urban land that is forecast for the coming decades—much of which is expected to occur in low-and middle-income countries with existing vulnerabilities to natural hazards[17]—thus has the potential to create growing hazardous interfaces for zoonotic pathogen exposure. In particular, the large effect sizes but sparser data availability for urban ecosystems (especially for mammals; Extended Data Fig. 4) highlight a key knowledge gap for anticipating the effects of urbanization on public health and biodiversity. Our findings support calls to enhance proactive human and animal surveillance within agricultural, pastoral and urbanizing ecosystems[33,34], and highlight the need to consider disease-related health costs in land use and conservation planning.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2562-8.

1. Myers, S. S. et al. Human health impacts of ecosystem alteration. *Proc. Natl Acad. Sci. USA* **110**, 18753–18760 (2013).
2. Gottdenker, N. L., Streicker, D. G., Faust, C. L. & Carroll, C. R. Anthropogenic land use change and infectious diseases: a review of the evidence. *EcoHealth* **11**, 619–632 (2014).
3. Keesing, F. et al. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
4. Ostfeld, R. S. & LoGiudice, K. Community disassembly, biodiversity loss, and the erosion of an ecosystem service. *Ecology* **84**, 1421–1427 (2003).
5. Johnson, P. T. J. et al. Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett.* **15**, 235–242 (2012).
6. Johnson, P. T. J., Preston, D. L., Hoverman, J. T. & Richgels, K. L. D. Biodiversity decreases disease through predictable changes in host community competence. *Nature* **494**, 230–233 (2013).
7. Newbold, T. et al. Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
8. Newbold, T. et al. Widespread winners and narrow-ranged losers: land use homogenizes biodiversity in local assemblages worldwide. *PLoS Biol.* **16**, e2006841 (2018).
9. Faust, C. L. et al. Pathogen spillover during land conversion. *Ecol. Lett.* **21**, 471–483 (2018).
10. Becker, D. J., Streicker, D. G. & Altizer, S. Using host species traits to understand the consequences of resource provisioning for host-parasite interactions. *J. Anim. Ecol.* **87**, 511–525 (2018).
11. Plowright, R. K. et al. Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
12. Shah, H. A., Huxley, P., Elmes, J. & Murray, K. A. Agricultural land-uses consistently exacerbate infectious disease risks in Southeast Asia. *Nat. Commun.* **10**, 4299 (2019).
13. Gottdenker, N. L., Chaves, L. F., Calzada, J. E., Saldaña, A. & Carroll, C. R. Host life history strategy, species diversity, and habitat influence *Trypanosoma cruzi* vector infection in changing landscapes. *PLoS Negl. Trop. Dis.* **6**, e1884 (2012).
14. Fornace, K. M. et al. Association between landscape factors and spatial patterns of *Plasmodium knowlesi* infections in Sabah, Malaysia. *Emerg. Infect. Dis.* **22**, 201–209 (2016).
15. Pulliam, J. R. C. et al. Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* **9**, 89–101 (2012).
16. Kilpatrick, A. M. Globalization, land use, and the invasion of West Nile virus. *Science* **334**, 323–327 (2011).
17. Popp, A. et al. Land-use futures in the shared socio-economic pathways. *Glob. Environ. Change* **42**, 331–345 (2017).
18. Civitello, D. J. et al. Biodiversity inhibits parasites: broad evidence for the dilution effect. *Proc. Natl Acad. Sci. USA* **112**, 8667–8671 (2015).
19. LoGiudice, K., Ostfeld, R. S., Schmidt, K. A. & Keesing, F. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proc. Natl Acad. Sci. USA* **100**, 567–571 (2003).
20. Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. Biol. Soc* **267**, 1947–1952 (2000).
21. Johnson, P. T. J., Ostfeld, R. S. & Keesing, F. Frontiers in research on biodiversity and disease. *Ecol. Lett.* **18**, 1119–1133 (2015).
22. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
23. Young, H., Griffin, R. H., Wood, C. L. & Nunn, C. L. Does habitat disturbance increase infectious disease risk for primates? *Ecol. Lett.* **16**, 656–663 (2013).
24. Hudson, L. N. et al. The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecol. Evol.* **7**, 145–188 (2017).
25. Lloyd-Smith, J. O. et al. Epidemic dynamics at the human–animal interface. *Science* **326**, 1362–1367 (2009).
26. Joseph, M. B., Mihaljevic, J. R., Orlofske, S. A. & Paull, S. H. Does life history mediate changing disease risk when communities disassemble? *Ecol. Lett.* **16**, 1405–1412 (2013).
27. Kamiya, T., O'Dwyer, K., Nakagawa, S. & Poulin, R. What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biol. Rev. Camb. Philos. Soc.* **89**, 123–134 (2014).
28. Lee, K. A., Wikelski, M., Robinson, W. D., Robinson, T. R. & Klasing, K. C. Constitutive immune defences correlate with life-history variables in tropical birds. *J. Anim. Ecol.* **77**, 356–363 (2008).
29. Rohr, J. R. et al. Towards common ground in the biodiversity–disease debate. *Nat. Ecol. Evol.* **4**, 24–33 (2020).
30. Hosseini, P. et al. Does the impact of biodiversity differ between emerging and endemic pathogens? The need to separate the concepts of hazard and risk. *Philos. Trans. R. Soc. Lond. B* **372**, 20160129 (2017).
31. Brearley, G. et al. Wildlife disease prevalence in human-modified landscapes. *Biol. Rev. Camb. Philos. Soc.* **88**, 427–442 (2013).
32. Burkett-Cadena, N. D. & Vittor, A. Y. Deforestation and vector-borne disease: Forest conversion favors important mosquito vectors of human pathogens. *Basic Appl. Ecol.* **26**, 101–110 (2018).
33. Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and disease emergence: dynamics at the wildlife–livestock-human interface. *Trends Ecol. Evol.* **32**, 55–67 (2017).
34. Holmes, E. C., Rambaut, A. & Andersen, K. G. Pandemics: spend on surveillance, not prediction. *Nature* **558**, 180–182 (2018).

Methods

We combined a global database of ecological assemblages (PREDICTS)[24] with data on host–pathogen and host–parasite associations, to create a global, spatially explicit dataset of local zoonotic host diversity. We define pathogens and parasites (henceforth 'pathogens') as including bacteria, viruses, protozoa, helminths and fungi (excluding ectoparasites). PREDICTS contains species records compiled from 666 published studies that sampled local biodiversity across land use type and intensity gradients, allowing global space-for-time analysis of land use effects on local species assemblages (that is, comparison between sites with natural vegetation considered to be a baseline). We analysed relative differences in wildlife host community metrics (zoonotic host species richness and abundance) between undisturbed (primary) land and nearby sites under varying degrees of anthropogenic disturbance. We subsequently conducted further analyses to examine how host species responses to land use vary across different mammalian and avian orders, and to test whether mammal pathogen richness (including both human and non-human pathogens) covaries with tolerance to land use.

Datasets

Ecological community and land use data. Each of the more than 3.2 million records in PREDICTS is a per-species, per-site measure of either occurrence (including absences) or abundance, alongside metadata on site location, land use type and use intensity. The database provides as representative a sample as possible of local biodiversity responses to human pressure, containing 47,000 species in a taxonomic distribution broadly proportional to the numbers of described species in major terrestrial taxonomic groups[24]. We first pre-processed PREDICTS following previous studies[7]: records collected during multiple sampling events at one survey site (for example, multiple transects) were combined into a single site record, and for studies for which the methods were sensitive to sampling effort (for example, area sampled), species abundances were adjusted to standardize sampling effort across all sites within each study, by assuming a linear relationship between sampling effort and recorded abundance measures (both following ref. [7]). Our analyses of species occurrence and richness are therefore based on discrete count data, whereas abundances are pseudo-continuous (counts adjusted for survey effort). Owing to the multi-source structure of PREDICTS (multiple studies with differing methods and scope), the absolute species richness and abundance measures are non-comparable between studies[24], so our analyses necessarily measure relative differences across land use classes.

Host–pathogen association data. We compiled animal host–pathogen associations from several source databases, to provide as comprehensive a dataset as possible of zoonotic host species and their pathogens: the Enhanced Infectious Diseases (EID2) database[35]; the Global Mammal Parasite Database v.2.0 (GMPD2) which collates records of parasites of cetartiodactyls, carnivores and primates[36]; a reservoir hosts database[37]; a mammal–virus associations database[22]; and a rodent zoonotic reservoirs database[38] augmented with pathogen data from the Global Infectious Disease and Epidemiology Network (GIDEON) (Supplementary Table 8). We harmonized species names across all databases, excluding instances in which either hosts or pathogens could not be classified to species level. To prevent erroneous matches due to misspelling or taxonomic revision, all host species synonyms were accessed from Catalogue Of Life using 'taxize' v.0.8.9[39]. Combined, the dataset contained 20,382 associations between 3,883 animal host species and 5,694 pathogen species.

Each source database applies different methods and taxonomic scope. EID2 defines associations broadly, on the basis of evidence of a cargo species being found in association with a carrier (host) species, rather than strict evidence of a pathogenic relationship or reservoir status[35]. The other four databases were developed using targeted searches of literature and/or surveillance reports, focus mainly on mammals, and provide more specific information on strength of evidence for host status (either serology, pathogen detection/isolation, and/or evidence of acting as reservoir for cross-species transmission). We therefore harmonized definitions of host–pathogen associations across the full combined database. Across all animal taxa we broadly defined associations on the basis of any documented evidence (cargo-carrier or stronger; that is, including all datasets). Additionally, for mammals only (owing to more comprehensive pathogen data availability), we were able to define two further tiers based on progressively stronger evidence: first, serological or stronger evidence of infection; and second, either direct pathogen detection, isolation or reservoir status. Across all pathogens, we also harmonized definitions of zoonotic status. Each pathogen was classified as human-shared if it was recorded as infecting humans within either one of the source host–pathogen databases or an external human pathogens list collated from multiple sources (Supplementary Table 8). Because the source datasets contain some organisms that infect humans and animals rarely or opportunistically, or that may not strictly be zoonotic (for example, pathogens with an environmental or anthroponotic reservoir), pathogens were also more specifically defined as zoonotic agents (aetiological agent of a specific human disease with a known animal reservoir) if classed as such in GIDEON, the Atlas of Human Infectious Diseases[40] or an additional human pathogens database[41].

Combined datasets of hosts and land use. We combined PREDICTS with the compiled host–pathogen database by matching records by species binomial, and each species record was given a binary classification of 'host' or 'non-host' of human-shared pathogens. We adopted a two-tiered definition of host status, to examine the effect of making more or less conservative assumptions about the likelihood of a species contributing to pathogen transmission dynamics and spillover to humans. First, we defined host status broadly: as any species with an association with at least one human-shared pathogen (as defined above), which for mammals must be based on serological or stronger evidence of infection (henceforth referred to as the 'full dataset'). 177 studies in PREDICTS contained host species matches (190 mammals, 146 birds, 1 reptile, 2 amphibians, 37 invertebrates; listed in Supplementary Table 1). Second, because mammals are the predominant reservoirs of both endemic and emerging zoonotic infections owing to their phylogenetic proximity to humans[42,43], we also defined mammal species as zoonotic reservoir hosts on the basis of stricter criteria: an association with at least one zoonotic agent (as defined above) that must be based on direct pathogen detection, isolation or confirmed reservoir status (henceforth referred to as 'mammal reservoirs subset'). Within PREDICTS, 63 studies contained host matches based on this narrower definition (143 mammal reservoir hosts; Extended Data Fig. 4, Supplementary Table 1).

Before analysis, we filtered PREDICTS to include only studies that sampled taxa relevant to zoonotic transmission, because the full database includes many studies with a different taxonomic scope (for example, plants or non-vector invertebrates)[24]. We retained all studies that sampled any mammal or bird species, as these groups are the main reservoir hosts of zoonoses. For all other taxa, given that zoonoses and their hosts occur globally, we made the more conservative assumption that studies with no sampled hosts represent false absences (that is, resulting from study aims and methodology) rather than true absences (that is, no hosts are present), and included only studies with at least one host match in one sampled site in community models. This resulted in a final dataset of 530,161 records from 6,801 sites in 184 studies (full dataset) and 51,801 records from 2,066 sites within 66 studies (mammal reservoirs dataset; including mammal studies only) (Fig. 1). Some host records were of arthropod vectors, but as these are a small proportion of records (around 2%; Supplementary Table 1) we generically refer to all matched species as 'hosts'. By matching on species binomial we assume that pathogens are equally likely to occur anywhere within their hosts' geographical range; evidence from terrestrial mammal orders suggests that this assumption is reasonable globally[44,45]. Although overlooking geographical variation in pathogen occurrence, pathogen geographical distributions are poorly understood and subject to change, making it difficult to define geographical constraints on host status.

We aggregated land use classes in PREDICTS to ensure a more even distribution of sampled sites. We assigned each survey site's land use type to one of four categories: primary vegetation, secondary vegetation, managed ecosystems (plantation forest, pasture and cropland) and urban. Land use intensity was assigned to either minimal, substantial (combining light and intense use) or cannot decide (the latter were excluded from models). Original use intensity definitions[7] reflect gradation of potential human effects within land use types; for example urban sites range from minimal (villages, large managed green spaces) to high intensity (impervious with few green areas). Land use categories simplify complex landscape processes, so our aggregation might mask subtle differences in disturbance mode and intensity. However, although some local studies have found differences in zoonotic host abundance and pathogen prevalence between different management regimes[46], we had no a priori reason to hypothesize differences between managed ecosystem types globally. Study regions were categorized as temperate or tropical, following ref. [47].

Statistical analysis

Accounting for species-level differences in pathogen discovery effort. The probability of identifying zoonotic pathogens within a species is strongly influenced by effort, meaning that poorly studied species in our data could be falsely classified as non-hosts. Because research effort might also positively correlate with species' abundance in anthropogenic landscapes, accounting for this uncertainty is crucial. In statistical models we therefore consider host status (and derived metrics such as host richness) to be an uncertain variable, by assuming that all known hosts in our dataset are true hosts (true positives), and that non-hosts comprise a mixture of true non-hosts and an unknown number of misclassified species. We propagate this uncertainty into all model estimates using a bootstrapping approach, in which each iteration transitions a proportion of non-host species to host status with a probability influenced by research effort and taxonomic group (with poorly researched species in taxonomic orders known to host more zoonoses having the highest transition rates; Extended Data Fig. 2, Supplementary Methods 1).

We estimate disease-related research effort using species publication counts extracted from the PubMed biomedical database (1950–2018) for every species within our dataset ($n = 7{,}285$; Extended Data Fig. 2c), following other studies in disease macroecology in which publication effort often explains much of the variation in response variables[22,48]. Across 100 randomly sampled mammal species from PREDICTS, PubMed publication counts were highly correlated to those from Web of Science and Google Scholar (both Pearson's $r = 0.93$), indicating robustness to choice of publications database. Using publication counts directly to index species misclassification probability is problematic, because the relationship between publication effort and host status is both nonlinear (for example, due to positive feedback, in which pathogen detection drives increasing research towards a species or taxon) and taxon-specific (for example, because some taxa are more intensely targeted for surveillance). We therefore calculate a trait-free approximation of false classification probability for non-host species (detailed in Supplementary Methods 1) by assuming, first, that the relative likelihood of a species being a zoonotic host is proportional to the number of known hosts in the same taxonomic order (that is, a poorly studied primate is more likely to be a zoonotic host than a poorly studied moth), and second, that confidence in non-host status accrues and saturates with increasing publication effort (following the cumulative curve of publication effort for known hosts within the same order; Extended Data Fig. 2a, b). Therefore, under-researched mammals, followed by birds, have the highest estimated false classification probabilities, but with substantial variation among mammalian and avian orders (Extended Data Fig. 2d, e).

Because data constraints prevent direct observation of how host detections accrue with discovery effort, our trait-free approximation leverages current knowledge of the distribution of zoonotic hosts and publication effort across broad taxonomic groups, and thus might over- or underestimate absolute host potential in any particular species. For example, because species traits and research effort are autocorrelated, our assumption that all non-host species per taxonomic group are equally likely to host zoonoses may conservatively overestimate host potential in less-researched species: many ecological traits that make species more likely to be poorly studied (for example, lower population densities, smaller range sizes[49,50]) would often be expected to reduce their relative importance in multi-host pathogen systems[51]. Nonetheless, our approach is sufficient to address the main confounding factor of our study—that is, the potential for biased distribution of research across land use types and biomes globally.

Community models of host species richness and total abundance. All modelling was conducted using mixed-effects regression in a Bayesian inference framework (Integrated Nested Laplace Approximation; INLA)[52]. We aggregated ecological communities data to site-level by calculating the per-site species richness (number of species) and total abundance (total number of sampled individuals, adjusted for survey effort) of host and non-host species. Land use type and intensity were combined into a categorical variable with 8 factor levels (type + intensity, for 4 types and 2 intensity levels). During model selection we considered fixed effects for land use and log-transformed 2005 human population density extracted from the Centre for International Earth Science Information Network (CIESIN) (because synanthropic species diversity might respond to changes in human population density independently of land use; Extended Data Fig. 8). All models included random intercept for study to account for between-study variation, and we additionally considered random intercepts for spatial block within study (to account for the local spatial arrangement of sites), site ID (to account for overdispersion caused by site-level differences)[7] and biome (as defined in PREDICTS).

We modelled the effects of land use on the richness and total abundance of host and non-host species separately, using a Poisson likelihood (log-link) to model species richness (discrete counts). Because abundance data were continuous after adjustment for survey effort, we followed other PREDICTS studies[7] and modelled log-transformed abundance with a Gaussian likelihood; log-transformation both reduces overdispersion and harmonizes interpretation of the fixed effects with the species richness models (that is, both measure relative changes in geometric mean diversity from primary land under minimal use). We also modelled the effects of land use on host richness and abundance as a proportion of overall site-level sampled species richness or abundance, by including log total species richness as an offset in Poisson models, and log total abundance as a continuous fixed effect (effectively an offset) in abundance models.

For each response variable we first selected among candidate model structures, comparing all combinations of random effects with all fixed effects included, and subsequently comparing all possible fixed effects combinations using the best-fitting random effects structure. In all cases we selected among models using the Bayesian pointwise diagnostic metric Watanabe-Akaike Information Criterion (WAIC)[53] (Supplementary Tables 3, 4). The final models were subsequently checked for fit and adherence to model assumptions, including testing for spatial autocorrelation in residuals (Extended Data Fig. 9). We then bootstrapped each final model for 1,000 iterations to incorporate research effort. For each iteration, each non-host species was randomly transitioned to host status as a Bernoulli trial with success probability p equal to estimated false classification probability (as described above; Supplementary Methods 1, Extended Data Fig. 2), all community response variables were recalculated, the model was fitted and 2,500 samples were drawn from the approximated joint posterior distribution. We then calculated posterior marginal parameter estimates (median and quantile ranges) across all samples from the bootstrap ensemble (Fig. 2, Supplementary Table 5). Between 90 and 150 non-host species (median 121) were selected to transition per iteration, increasing the total number of hosts by 24–40% (median 32%; Extended Data Fig. 2e). Because study coverage is heterogeneous globally, we subjected the full model ensembles to random and geographical cross-validation (Extended Data Fig. 3). We also conducted the same modelling procedure using only the strictly defined mammal reservoirs subset (Extended Data Fig. 4).

Species-level estimates of land use effects on mammalian and avian zoonotic hosts. Because aggregate community diversity metrics might mask important variation between taxonomic groups, we separately modelled the average effects of land use type on the occupancy and abundance of all hosts and non-hosts of zoonotic agents within five mammalian orders (Carnivora, Cetartiodactyla, Chiroptera, Primates, Rodentia) and two avian orders (Passeriformes, Psittaciformes). For mammals we defined zoonotic host status strictly (pathogen detection, isolation or confirmed reservoir status, as described above) and excluded urban sites owing to sparse urban sampling for mammals in PREDICTS (only 2 studies). All models included an interaction term between land use type and zoonotic host status (host or non-host) and random intercepts for each species–study combination and for taxonomic family (to account for gross phylogenetic differences). We again accounted for variable research effort per species as described above, fitting 500 models per order, and calculating posterior marginal estimates across samples drawn from the whole ensemble (Supplementary Table 6).

Abundance data were overdispersed and zero-inflated owing to the high proportion of absence records (that is, sites where species were not found despite being sampled for). We therefore used a hurdle-model-based approach[54] to estimate the effects of land use on abundance, by separately fitting occurrence models (presence-absence; binomial likelihood, logit-link) to the complete dataset for each mammalian order, and zero-truncated abundance models (ZTA, log-abundance with Gaussian likelihood) to the dataset with absences removed (Extended Data Fig. 5). Mean differences in abundance across land uses are then calculated as the product of the proportional differences in predicted occurrence probability and ZTA relative to primary land[54]. We used posterior samples from paired occurrence (transformed to probability scale) and ZTA models (transformed to linear scale) to calculate a distribution of hurdle predictions separately for each bootstrap iteration (that is, with the same non-hosts reclassified). We then summarized predicted changes per land use type across samples from the entire bootstrap ensemble (median and quantile ranges; Fig. 3). Owing to the complex nested structure of PREDICTS, our hurdle predictions assume independence between occurrence and ZTA processes, so do not formally account for the possibility of covariance at random effects (species or family) level. For clarity, we therefore show the contributions of each separate model for each order (Extended Data Fig. 5, Supplementary Table 6). In most orders, and when fitting models across all mammal species, land use often seems to act most consistently on species occurrence, with more variable effects on ZTA, suggesting that the independence assumption may be broadly reasonable at this global and cross-taxa scale.

Relationship between pathogen richness and responses to land use across mammal species. Pathogen richness (the number of pathogens hosted by a species) is a widely analysed trait in disease macroecology, with both overall pathogen richness, shared pathogen richness (that is, number of pathogens shared between focal species) and zoonotic pathogen richness often correlated to species traits such as intrinsic population density, life history strategy and geographic range size[5,22,27,55]. If human-disturbed landscapes systematically select for species trait profiles that facilitate host status, we might expect to observe positive responses to land use in species with higher richness of either human-shared or non-human-shared pathogens. We tested this hypothesis for mammals, owing to availability of much more comprehensive pathogen data than for other taxa, by analysing the relationship between species pathogen richness and probability of occurrence across three land use types (primary, secondary and managed; urban sites excluded owing to limited sampling).

Within the subset of PREDICTS studies that sampled for mammals, containing 26,569 records of 546 mammal species (1950 sites, 66 studies), we used the host–pathogen association dataset to calculate, first, each mammal species' richness of human-shared pathogens, and second its richness of pathogens with no evidence of infecting either humans or domestic animals ('non-human-shared'), defining associations on the basis of serological evidence or stronger. Of the 546 mammals, 190 species had at least one known human-shared pathogen (human-shared pathogen richness mean 1.92, s.d. 6.07) and 96 species had at least one non-human-shared pathogen (non-human-shared pathogen richness mean 0.81, s.d. 4.16). We account for research effort differently than in the binary host status models above, because pathogen richness is a continuous variable that is influenced by magnitude of effort (that is, more effort would be expected to increase the number of detected pathogens; Extended Data Fig. 6b, c). Therefore, we account for effort by estimating per-species residual pathogen richness not explained by publication effort (that is, the difference between observed pathogen richness and expected pathogen richness given publication effort and taxonomic group). To do this, we modelled the effect of publication effort on pathogen richness (discrete counts) separately for human-shared and non-human-shared pathogens, using a Poisson likelihood with a continuous fixed effect of log-publications and random intercepts and slopes for each mammalian order and family (to account for broad taxonomic differences in host–pathogen ecology between orders[22]). We fitted the model to data from all mammal species in our host–pathogen database ($n = 780$) and predicted expected mean pathogen richness for all mammals in PREDICTS. We calculated residuals from observed values for these species (Extended Data Fig. 6), which we expect represent trait-mediated variation, given the evidence that mammal pathogen richness covaries with species traits after accounting for phylogeny and research effort[22].

We then modelled the relationship between residual pathogen richness (scaled to mean 0, s.d. 1) and species probability of occurrence across land use types, separately for human-shared and non-human-shared pathogens (Extended Data Fig. 7). Species occurrence was modelled using a binomial (logit-link) likelihood, with fixed effects for the interaction between residual pathogen richness and land use type, and random intercepts for species, order, study and spatial block within study. As with previous analyses, models were checked for fit and adherence to assumptions. Pathogen surveillance in animals is often focused on species of zoonotic concern, meaning that pathogen inventories (especially of non-human-shared pathogens) may be more complete for some taxonomic groups than others. We therefore tested model sensitivity to separately fitting models containing, first, only species from the four most comprehensively sampled mammalian orders for parasites and pathogens (Primates, Cetartiodactyla, Perissodactyla and Carnivora; the focal taxa of the Global Mammal Parasite Database[36]), and second, species from all other mammal orders. We also tested for sensitivity to uncertainty in the publications–pathogen richness relationship, by separately fitting the land use model to 400 sets of residuals derived using posterior samples from the fitted publication effort model (Extended Data Fig. 6g, h), and summarizing parameters across the full ensemble. Fixed effects directions and strength of evidence were consistent across all models (Supplementary Table 7). Data processing and analyses were conducted in R v.3.4.1[56], with model inference conducted in R-INLA[52].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data sources are listed, with links to freely available online sources, in Supplementary Table 8. Where not freely available online, all data for this study are archived at Figshare https://doi.org/10.6084/m9.figshare.7624289. Source data are provided with this paper.

Code availability

All code for this study is available at Figshare https://doi.org/10.6084/m9.figshare.7624289.

35. Wardeh, M. et al. Database of host-pathogen and related species interactions, and their global distribution. *Sci. Data* **2**, 150049 (2015).
36. Stephens, P. R. et al. Global Mammal Parasite Database version 2.0. *Ecology* **98**, 1476 (2017).
37. Plourde, B. T. et al. Are disease reservoirs special? Taxonomic and life history characteristics. *PLoS ONE* **12**, e0180716 (2017).
38. Han, B. A., Schmidt, J. P., Bowden, S. E. & Drake, J. M. Rodent reservoirs of future zoonotic diseases. *Proc. Natl Acad. Sci. USA* **112**, 7039–7044 (2015).
39. Chamberlain, S. & Szocs, E. Taxize: taxonomic search and retrival in R. *F1000Res* **18**, 191 (2013).
40. Wertheim, H. F. L., Horby, P. & Woodall, J. P. *Atlas of Human Infectious Diseases* (Wiley-Blackwell, 2012).
41. Taylor, L. H., Latham, S. M. & Woolhouse, M. Risk factors for human disease emergence. *Philos. Trans. R. Soc. B* **356**, 983–989 (2001).
42. Han, B. A., Kramer, A. M. & Drake, J. M. Global patterns of zoonotic disease in mammals. *Trends Parasitol*. **32**, 565–577 (2016).
43. Røttingen, J. A. et al. New vaccines against epidemic infectious diseases. *N. Engl. J. Med.* **376**, 610–613 (2017).
44. Harris, N. C. & Dunn, R. R. Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. *Ecol. Lett*. **13**, 1411–1418 (2010).
45. Cooper, N. et al. Phylogenetic host specificity and understanding parasite sharing in primates. *Ecol. Lett*. **15**, 1370–1377 (2012).
46. Young, H. S. et al. Interacting effects of land use and climate on rodent-borne pathogens in central Kenya. *Philos. Trans. R. Soc. B* **372**, 20160116 (2017).
47. Newbold, T. et al. Global patterns of terrestrial assemblage turnover within and among land uses. *Ecography* **39**, 1151–1163 (2016).
48. Cooper, N., Kamilar, J. M. & Nunn, C. L. Host longevity and parasite species richness in mammals. *PLoS ONE* **7**, e42190 (2012).
49. González-Suárez, M., Lucas, P. M. & Revilla, E. Biases in comparative analyses of extinction risk: mind the gap. *J. Anim. Ecol*. **81**, 1211–1222 (2012).
50. Ducatez, S. & Lefebvre, L. Patterns of research effort in birds. *PLoS ONE* **9**, e89955 (2014).
51. Paull, S. H. et al. From superspreaders to disease hotspots: linking transmission across hosts and space. *Front. Ecol. Environ*. **10**, 75–82 (2012).
52. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. Spatial and spatio-temporal models with R-INLA. *Spat. Spatio-Temporal Epidemiol*. **4**, 33–49 (2013).
53. Hooten, M. B. & Hobbs, N. T. A guide to Bayesian model selection for ecologists. *Ecol. Monogr*. **85**, 3–28 (2015).
54. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2010).
55. Dallas, T. A. et al. Host traits associated with species roles in parasite sharing networks. *Oikos* **128**, 23–32 (2019).
56. R: A language and environment for statistical computing. http://www.R-project.org/ (R Foundation for Statistical Computing, 2019).

Acknowledgements We thank L. Enright, A. Etard, L. Franklinos, R. Freeman, R. Lowe and R. Pearson for discussion on previous versions of the manuscript. This research was supported by a University College London Graduate Research Scholarship (R.G.); the Ecosystem Services for Poverty Alleviation Programme, Dynamic Drivers of Disease in Africa Consortium, NERC project no. NE-J001570-1 (D.W.R. and K.E.J.); an MRC UKRI/Rutherford Fellowship (MR/R02491X/1) and Wellcome Trust Institutional Strategic Support Fund (204841/Z/16/Z) (both to D.W.R.); and a Royal Society University Research Fellowship (T.N.). C.A.D. thanks the UK MRC and DFID for Centre funding (MR/R015600/1), and the UK National Institute for Health Research Health Protection Research Unit in Modelling Methodology at Imperial College London in partnership with Public Health England for funding (grant HPRU-2012–10080).

Author contributions R.G., D.W.R., K.E.J., T.N. and T.M.B. conceived and designed the study. C.A.D. contributed to the design of statistical analyses. R.G. collated and processed the data, and led and conducted the analyses with D.W.R., K.Q.C. and T.N. All authors contributed to writing the manuscript.

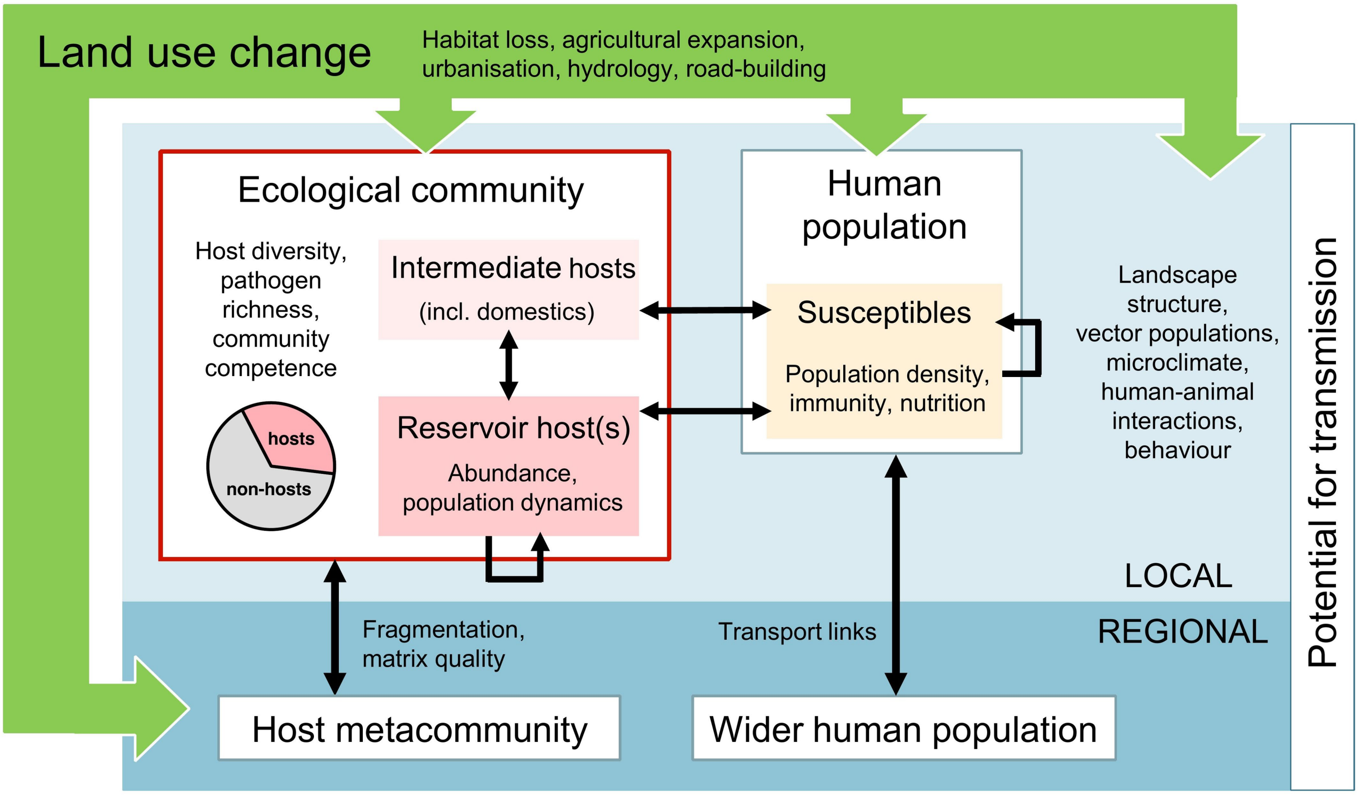
Competing interests The authors declare no competing interests.

Additional information

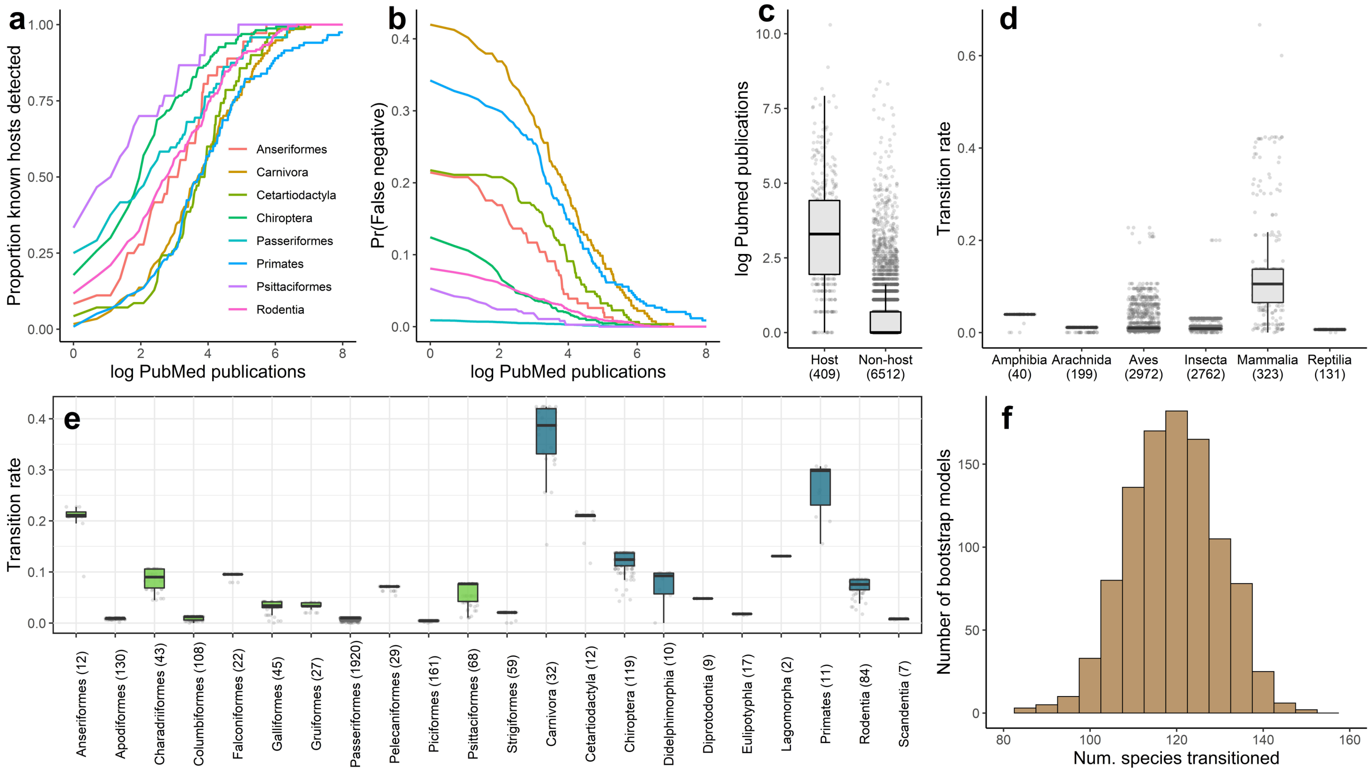
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2562-8.

Correspondence and requests for materials should be addressed to D.W.R. or K.E.J.

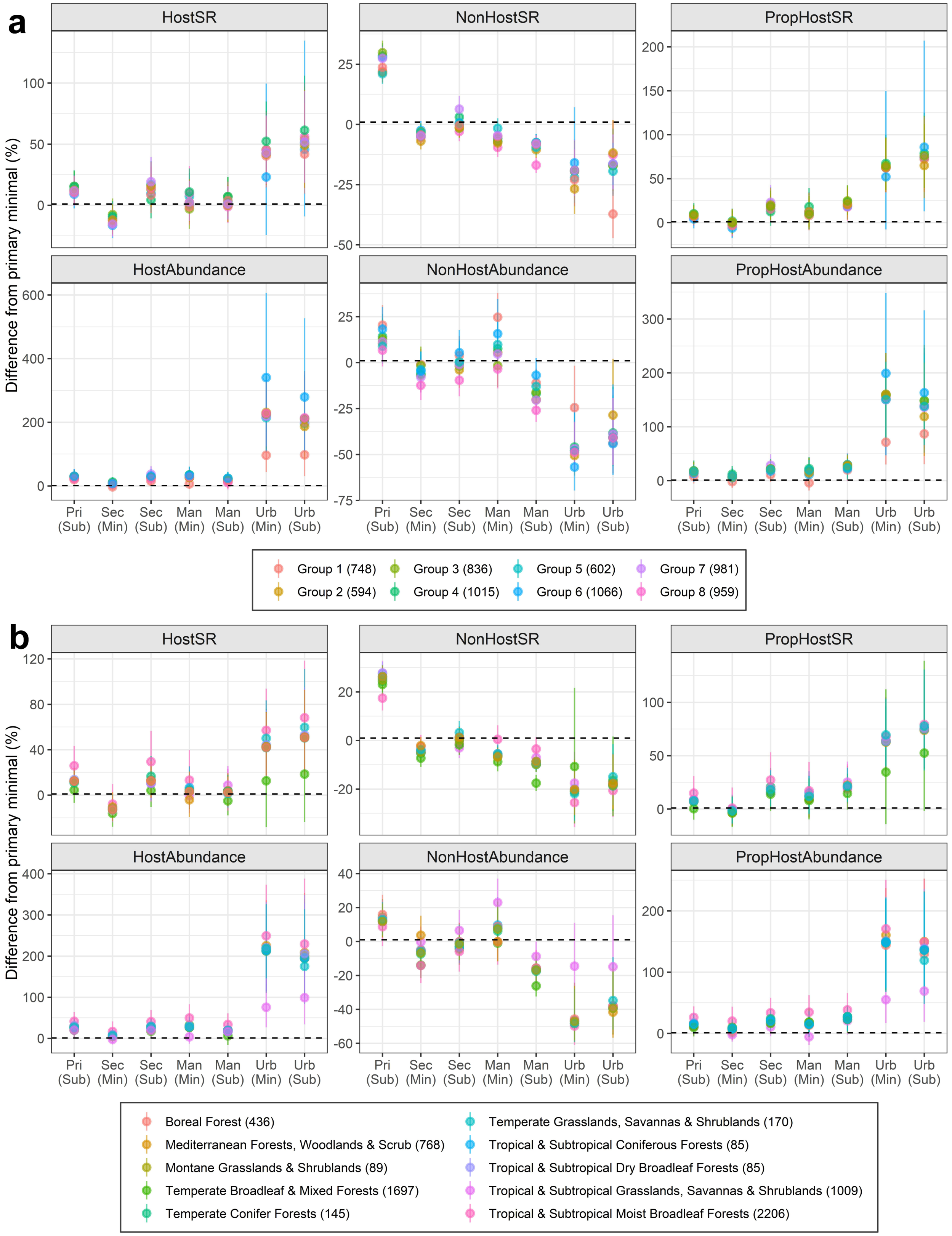
Peer review information *Nature* thanks Noam Ross and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Conceptual framework for the effects of land use change on zoonotic disease transmission. Pathogen transmission between potential hosts is shown as black arrows. Land use change (green driver) acts on ecological community composition and human populations (white boxes), and on environmental features that influence contact and transmission both locally (light blue box) and at broader geographical scales (dark blue box). These processes occur within a broader socio-ecological system context also influenced by additional environmental (for example, climatic), socioeconomic and demographic factors. Unpicking the relative influence of these different processes on disease outcomes is challenging in local disease system studies, in which multiple processes may be acting on pathogen prevalence and transmission intensity. The aim of this analysis was therefore to specifically examine, at a global scale, the effects of land use change on the composition of the potential host community (excluding domestic species), denoted by the red box.

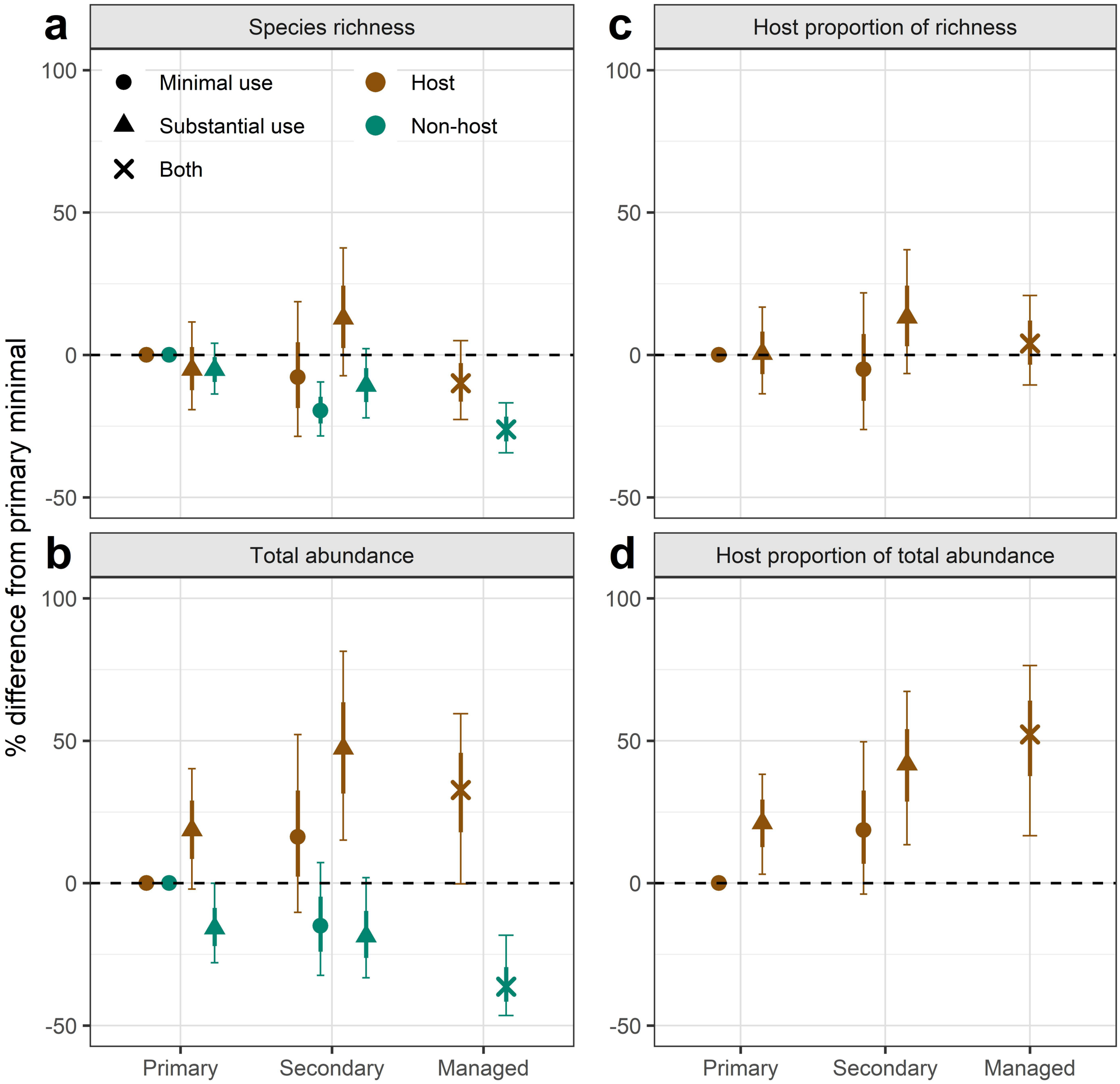


Extended Data Fig. 2 | Approximating research effort bias for non-host species within the PREDICTS dataset. For all non-host species, we approximated the likelihood of false classification given research effort (that is, probability of being a host, but not detected), based on the distribution of publication effort across known zoonotic hosts within the same taxonomic order (Supplementary Methods 1). **a**, **b**, Line graphs show, for several orders, the cumulative curve of publication counts for known zoonotic hosts (**a**; shown on log-scale), and approximated false classification probability, which declines and asymptotes with increasing levels of research effort (**b**) (line colours denote taxonomic order). **c**–**e**, Points and box plots show the distribution of PubMed publications for all host and non-host species in PREDICTS (**c**; total *n* = 6,921), and false classification probabilities (used as bootstrap transition rates) for all non-host species per taxonomic class in PREDICTS (**d**; total *n* = 3,665), and per key mammalian and avian order (**e**; total *n* = 2,927) (bracketed numbers denote number of species per group; boxes show median and interquartile range, whiskers show values within 1.5× IQR from quartile). **f**, The histogram shows the number of non-host species transitioned to host status for each of 1,000 bootstrapped models of the full dataset (median 121, 95% quantile range 102–142).

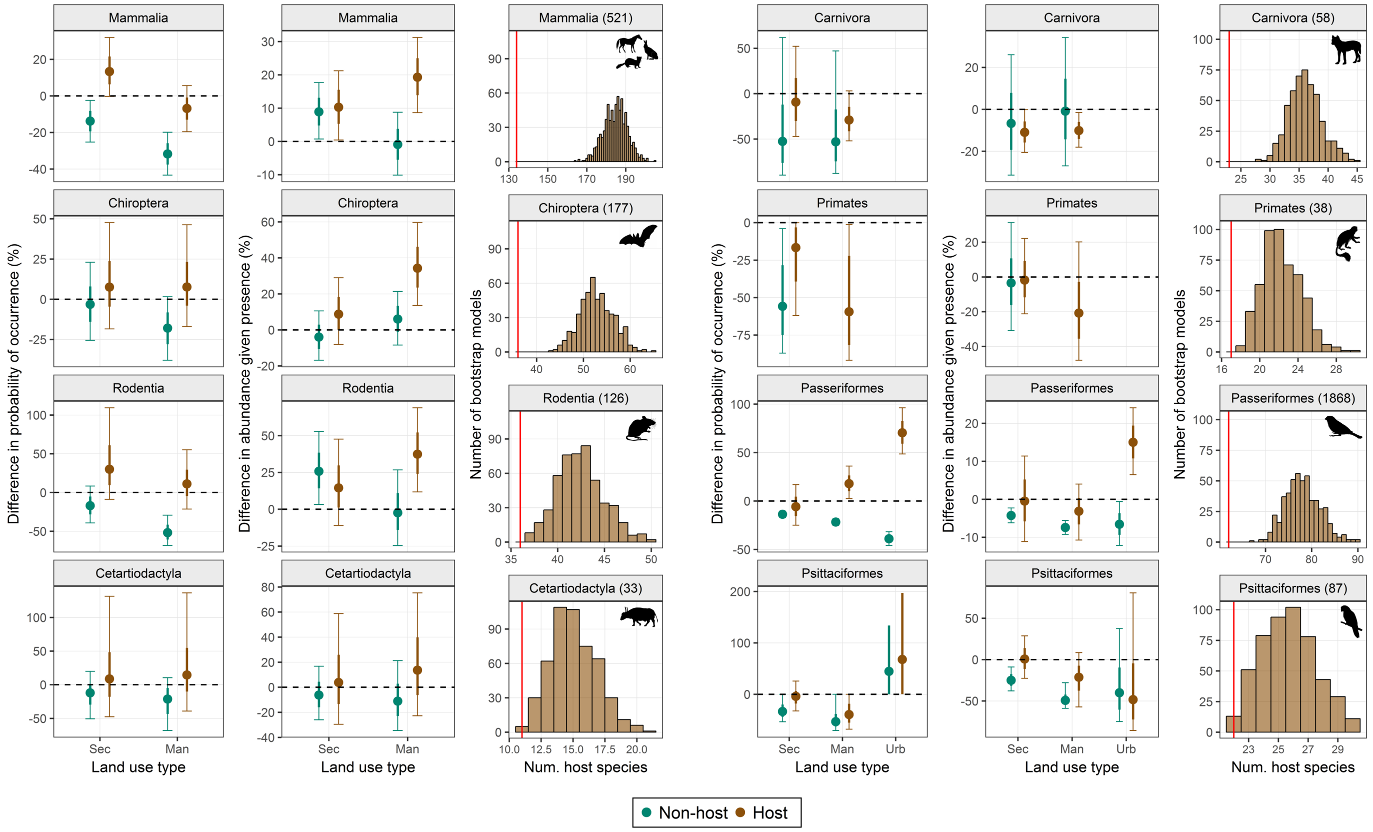


Extended Data Fig. 3 | See next page for caption.

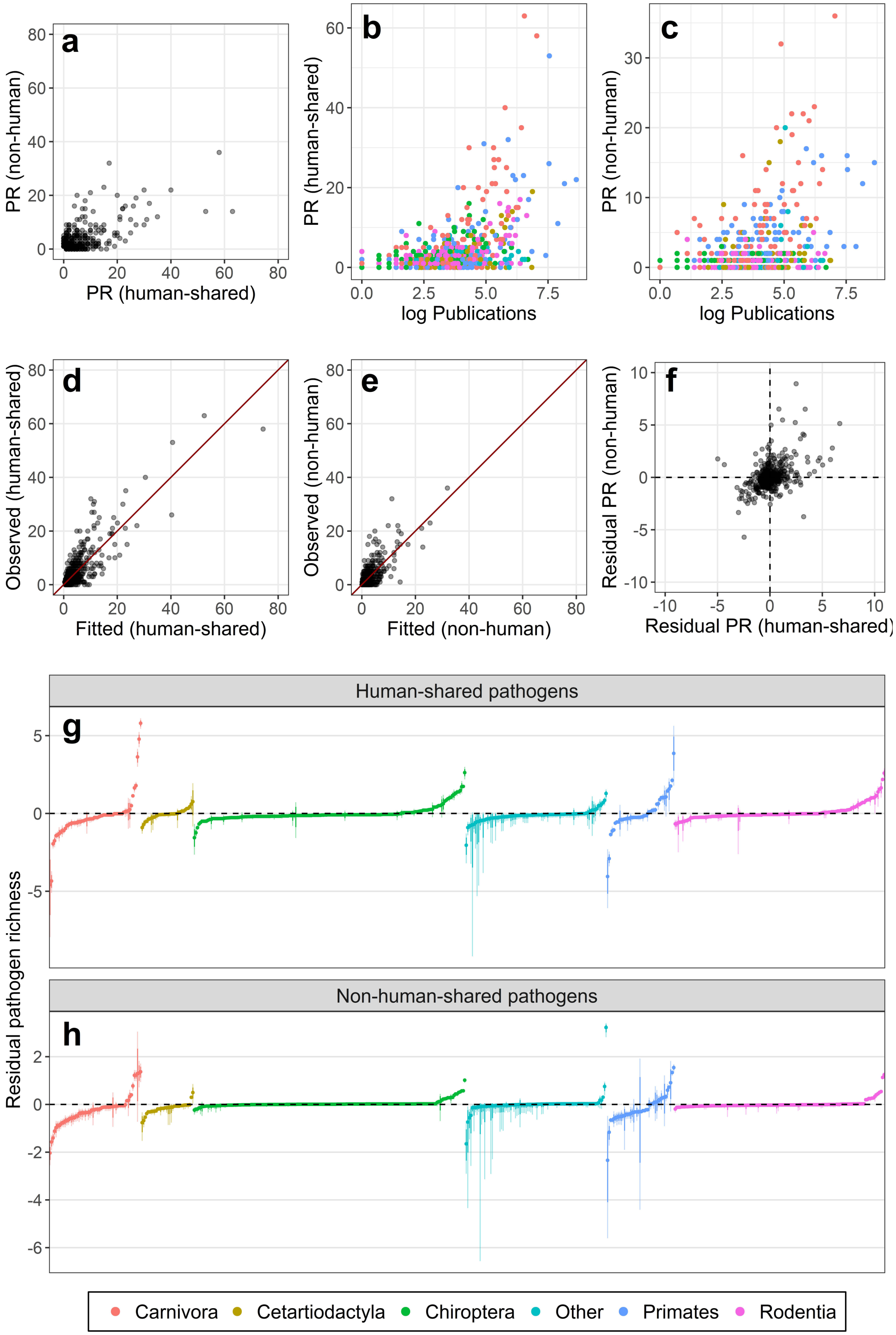
Extended Data Fig. 3 | Random (study-level) and geographical cross-validation of community models (full dataset). We tested the sensitivity of fixed effects estimates to both random and geographically structured (biome-level) subsampling. **a**, For random tests we fitted 8 hold-out models, excluding all sites from 12.5% of studies at a time (mean 12.5% of total sites excluded per model, range 4–19%). **b**, For geographical tests we fitted 14 hold-out models, with each excluding all sites from one biome (mean 7% of sites excluded per model, range 0.07–32%). Points and error bars show posterior marginal parameter distributions for each hold-out model (median and 95% quantile range, with colour denoting hold-out group or biome), calculated across samples from 500 bootstrap iterations per-model to account for variable research effort across species. Directionality and evidence for fixed-effects estimates are robust to both tests, suggesting that our results are not driven by data from any particular subset of studies or regions. Urban parameters are, however, the most sensitive to exclusion of data, probably owing to the relatively sparse representation of urban vertebrate diversity in the PREDICTS database (17 studies in our full dataset).



Extended Data Fig. 4 | Effects of land use on site-level mammalian reservoir host species richness and total abundance. a–**d**, Points, wide and narrow error bars show differences in diversity metrics from primary minimal use baseline (posterior marginal median, 67% and 95% quantile ranges respectively, across 1,000 bootstrap models). Models are of species richness (**a**) and total abundance (**b**) of reservoir host and all other (non-host) species, and of hosts as a proportion of site-level richness (**c**) and total abundance (**d**). For managed and urban sites, use intensities were combined to improve evenness of sampling (*n* = 2,026 sites from 63 studies: primary (589 and 572 for minimal and substantial use respectively), secondary (144, 257), managed (348) and urban (116)). Posterior estimates were calculated across an ensemble of 1,000 bootstrapped models (median 51, range 38–62 non-hosts transitioned to host status, that is, increasing host number by 28–46%) (Methods). Results from urban sites show the same trend as the full dataset (Fig. 2), but are not visualized owing to wide uncertainty: 88.7% (−2.1, 252.3) proportion richness, 307% (78.8, 500.7) proportion abundance (posterior median and 95% quantile range; see Supplementary Table 4). Point shape indicates use intensity (minimal, substantial or both combined) and colour indicates host (brown) or non-host (green). Reservoir species are listed in Supplementary Table 1 (mammal species listed as 'Detection/reservoir' in the 'Evidence of host status' column).

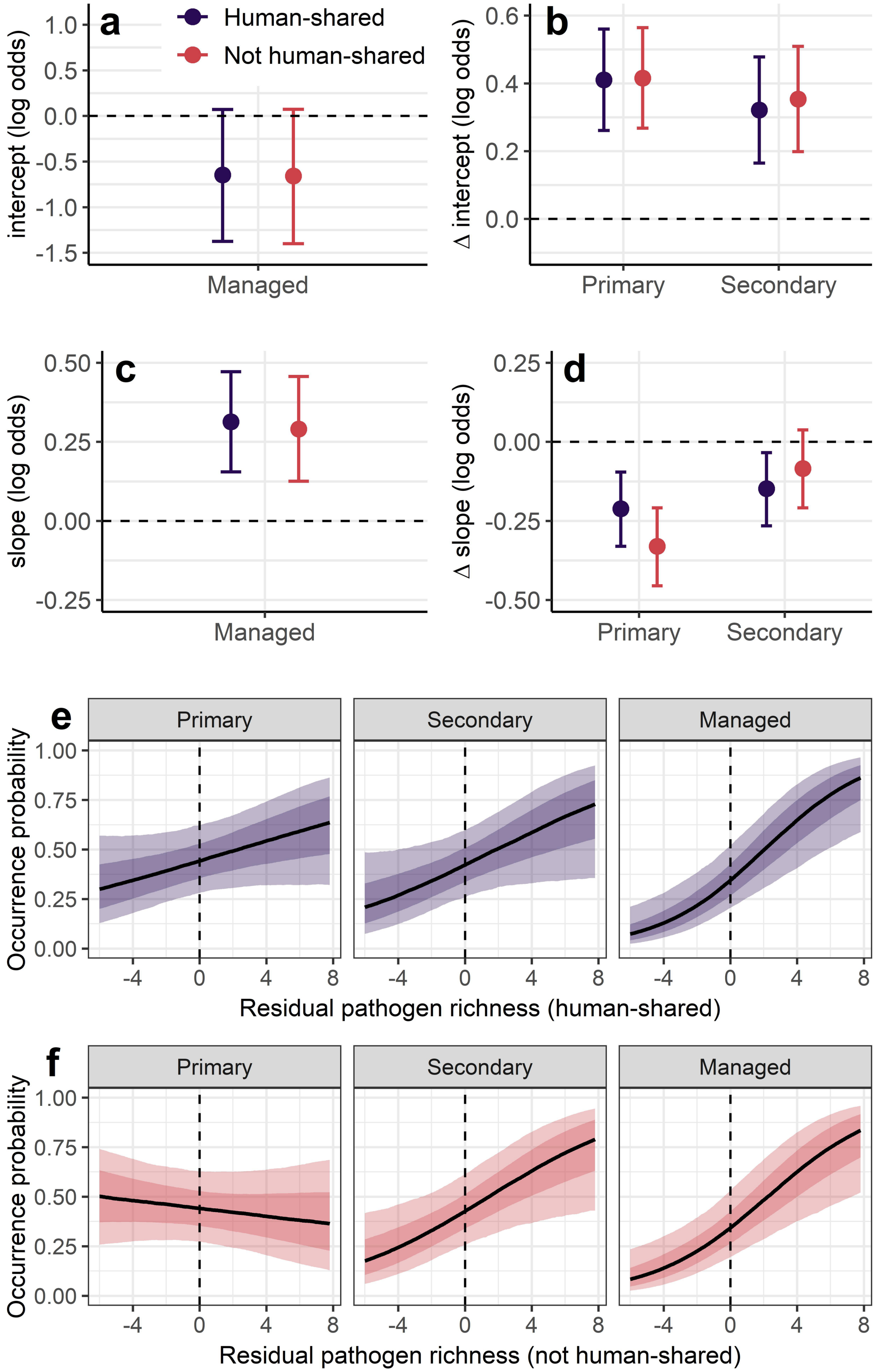


Extended Data Fig. 5 | Effects of land use on occurrence and zero-truncated abundance (abundance given presence) of mammalian and avian hosts and non-hosts of zoonotic agents. Each row of three plots shows the results of species-level modelling for each of five mammalian and two avian orders, and for mammals overall. Points, wide and narrow error bars show average difference in species occurrence probability (left column) and ZTA (middle column) (posterior median, 67% and 95% quantile ranges across 500 and 750 bootstrap iterations, for each order and all mammals respectively). Differences are shown in secondary (Sec), managed and urban sites relative to a primary land baseline (dashed line), across all host (brown) and non-host (green) species. Histograms show, for each taxonomic group, the distribution of host species counts across all bootstrap models (that is, after reclassifying non-hosts) compared to current number of known hosts (red vertical line), and the total number of species included in models (brackets in plot title). Estimates from occupancy and ZTA models (Supplementary Table 6) were combined, assuming independence of processes, to give the hurdle predictions in Fig. 3. Mammal reservoir status was defined on the basis of strict criteria (pathogen detection or isolation), and the full list of host species included in these estimates is provided in Supplementary Table 1 (scored '1' in the' zoonotic agent host' column). Silhouettes obtained from PhyloPic (http://phylopic.org/).



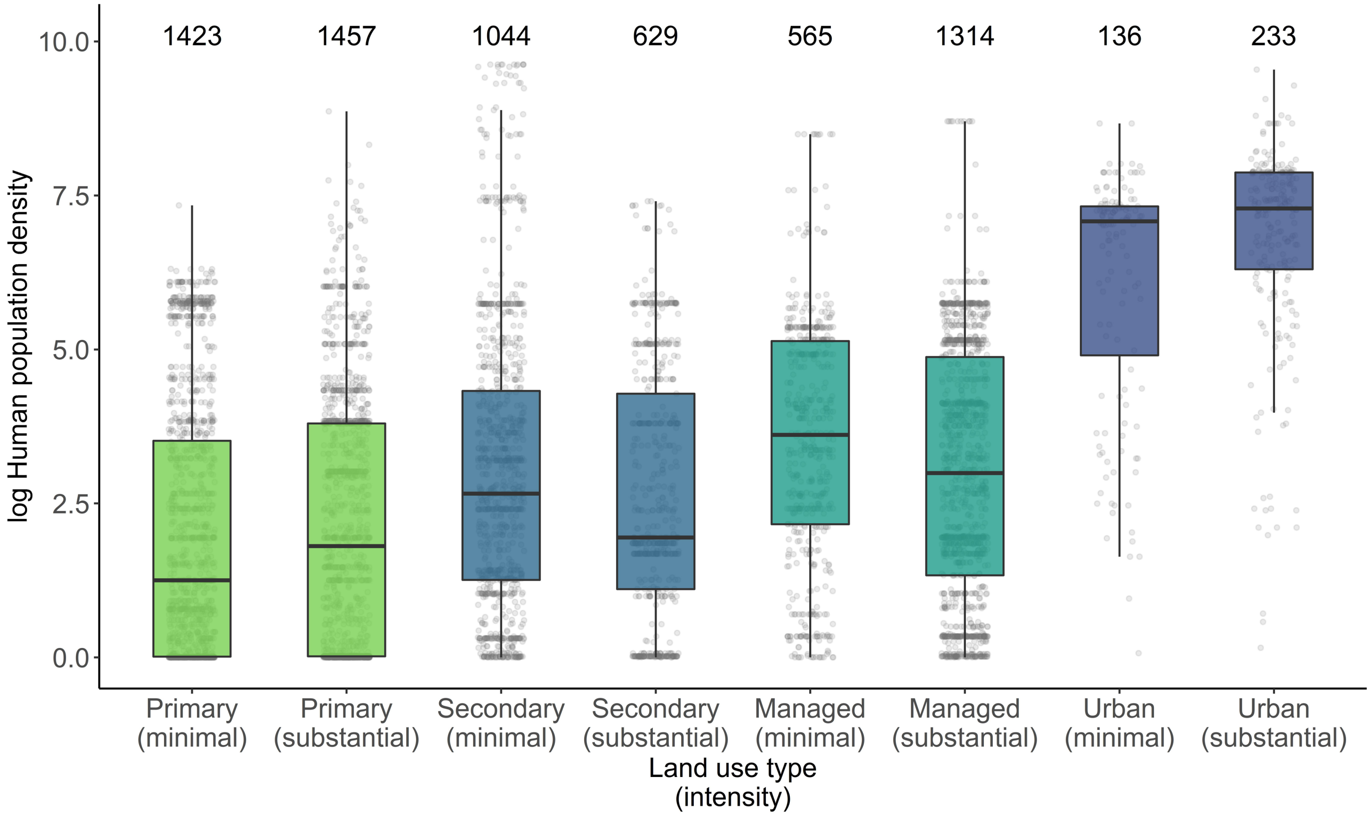
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Residual human-shared and non-human-shared pathogen richness across mammals. a–**c**, Distribution of human-shared and non-human-shared pathogen richness (**a**) and relationship to publication counts (**b**, **c**) are shown for mammals in our host–pathogen association dataset (n = 780 species; points represent species shaded by order, associations defined on serological or stronger evidence). **d**, **e**, Observed versus fitted plots show where observed deviates from expected pathogen richness given log-publications and taxonomic group (Poisson likelihood with random intercepts and slopes for order and family; slope estimates for log-publications are similar for both human and non-human-shared pathogens, β of 0.298 and 0.248 respectively). **f**, Fitted models were used to predict expected pathogen richness for mammals in PREDICTS (n = 546) and derive residuals from observed values, which were used in land use models (Extended Data Fig. 7). **g**, **h**, Calculating per-species residual quantile ranges across 2,500 posterior parameter samples shows that within-species residual variance is generally small relative to residual size, points and error-bars show posterior median, 67% and 95% intervals, scaled to unit variance), and land use model results are robust to including this uncertainty (Methods, Supplementary Table 7).

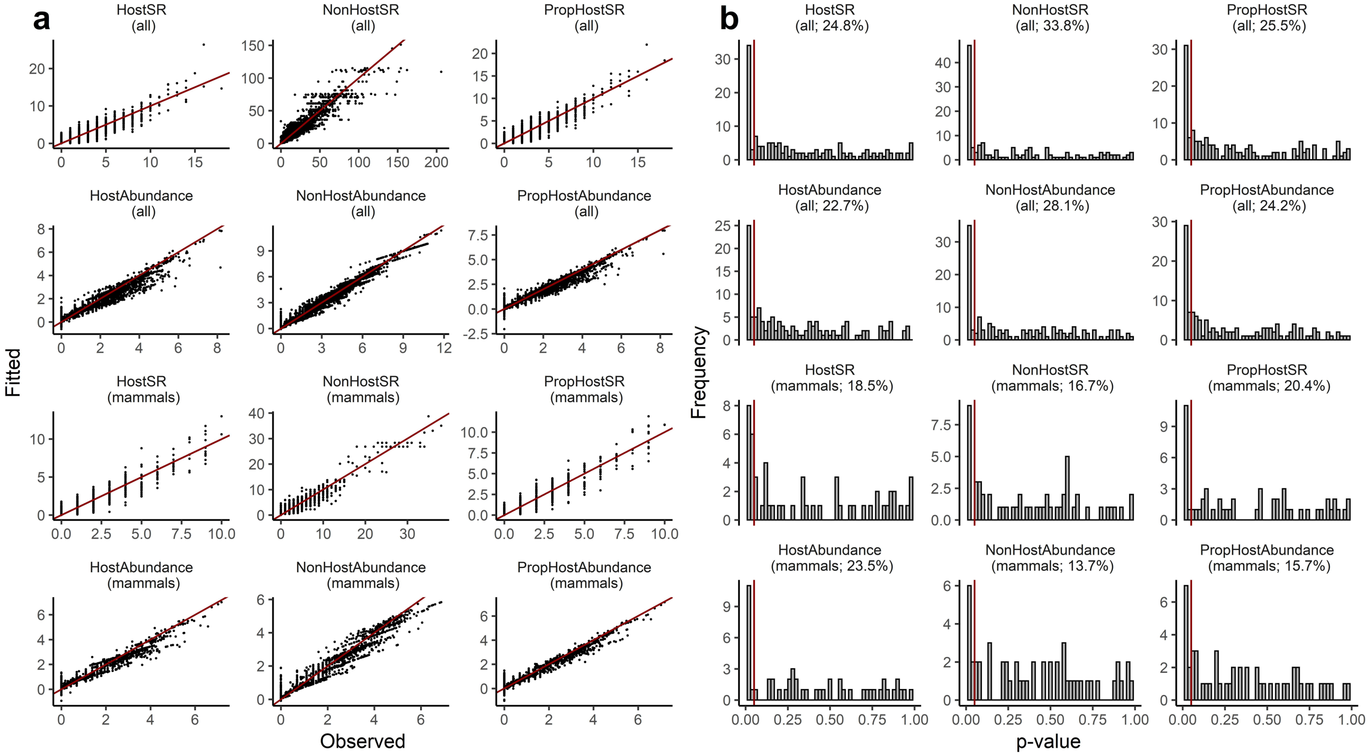


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Effects of land use on the relationship between mammal species pathogen richness and occurrence probability. **a**–**d**, Points and error bars show intercept (**a**, **b**) and slope parameters (**c**, **d**) of the relationship between residual pathogen richness (scaled to mean 0 and unit variance) and mammal species occurrence probability (on the log odds scale; median and 95% credible interval). Model was fitted to occurrence data for all mammals in the database (*n* = 29,569 records of 546 species, 1,950 sites, 66 studies). Intercept parameters represent the average occurrence probability of a species with residual pathogen richness of 0 (that is, with average pathogen richness given research effort and taxonomy), and slope parameters represent the change in occurrence probability for one scaled unit (s.d.) increase in residual pathogen richness (Extended Data Fig. 6g, h). Intercept and slope parameters for primary and secondary land measure the differences relative to managed land (that is, delta-intercept or delta-slope; **b**, **d**). **e**, **f**, Plotted lines show these relationships on the probability scale, showing the median (black line), 67% (dark shading) and 95% (light shading) quantile range, based on 3,000 samples from the joint posterior distribution. For both human-shared and non-human-shared pathogens, there is a positive relationship between the residual pathogen richness of a species and its probability of occurrence in human-managed land. For human-shared pathogens, the strength of this relationship (slope parameter) is significantly larger in managed sites than in both primary and secondary land, and for non-human-shared pathogens significantly larger in managed than in primary land (**d**; slopes for primary land not significantly different from 0). Full model summaries and results of sensitivity analyses are in Supplementary Table 7.



Extended Data Fig. 8 | Differences in human population density between land use types, for all sites within the full dataset. Points and boxplots show the distributions of log-transformed human population density by land use type and intensity, across all sites included in community models (*n* = 6,801). Boxes show median and interquartile range with whiskers showing values within 1.5 × IQR from quartile, and are coloured by land use type, and numbers denote the number of sites in each category. Human population density estimates were extracted from CIESIN Gridded Population of the World 4, for 2005, the median year of studies included in the dataset. Per-site log human density estimates were considered as fixed effects in community models of host diversity, because human-tolerant or synanthropic species might respond to human population change independently of land use (Methods).



Extended Data Fig. 9 | Diagnostic plots for all community models (full dataset and mammal reservoirs subset). Species richness counts were modelled with a Poisson likelihood, and abundance (adjusted counts) were log-transformed and modelled with a Gaussian likelihood (see Methods). Plot titles refer to model response variables: species richness (SR), total abundance (Abundance), for hosts, non-hosts, and for hosts as a proportion of the community (Prop). **a**, **b**, Observed data against model-fitted values are shown in **a**. The red line shows the expectation if observed equals fitted ($n = 6{,}801$ for full SR; $n = 6{,}093$ for full abundance; $n = 2{,}026$ for mammals SR; $n = 1{,}963$ for mammals abundance). We also tested for spatial autocorrelation of residuals across all sites within each study, with histograms (**b**) showing the distribution of per-study Moran's *I* *P* values (indicating significance of spatial autocorrelation among sites within that study) for each model ($n = 184$ for full SR; $n = 164$ for full abundance; $n = 63$ for mammals SR; $n = 60$ for mammals abundance). Numbers in brackets are the percentage of studies that contained significant spatial autocorrelation ($P < 0.05$, shown as a red line). Overall, spatial autocorrelation was fairly low across the dataset (statistically significant in 14–34% of studies, with maximum 26% for models with host metrics as response variables). Residuals and statistics were derived from a single fitted model including community mean false classification probability as a linear covariate to account for research effort (with known hosts given a false classification probability of 0), rather than the full bootstrap ensemble.

nature research

Corresponding author(s): Kate E. Jones, David W. Redding

Last updated by author(s): 20/05/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All data used in this study were collected and collated from freely-available open-source databases and the scientific literature. A full list of data sources, including citations and links, is provided in the Supplementary Information (Supp. Table 7). Any data that are not free-to-download are provided, along with code, in the accompanying Figshare repository (https://figshare.com/s/cede860bb9c77abf23ff)
Data analysis	All data processing and analyses were conducted using R v. 3.4.1. Bayesian mixed-effects models were specified and fitted using the package 'R-INLA'. All code used to conduct the analyses, and sufficient to reproduce results, is provided in the accompanying Figshare repository (https://figshare.com/s/cede860bb9c77abf23ff)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Code and data used in this analysis (where not freely available online) are archived at Figshare, and data sources are fully listed in Supplementary Table 8.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We combine a global database of local ecological communities data (site-level species occurrences/abundances) with global databases of species-level host-pathogen associations, to test the hypothesis that land use has predictable and positive effects on the richness and abundance of hosts of human parasites and pathogens. To do this, we model the effects of land use type and intensity (categorical independent variables) on host and non-host diversity metrics, comparing responses in disturbed sites to a minimally-disturbed (primary land) baseline, across 6801 sites from 184 published studies. We go on to analyse potentially important taxonomic variability in host species responses across mammals and birds by estimating average species-level differences in occurrence and abundance across land use types within important zoonotic host taxa. Lastly, we test for covariance between a species' overall pathogen richness (number of either human-shared or non human-shared pathogens) and its probability of occurring in human-disturbed landscapes. All analyses were conducted in a Bayesian hierarchical (mixed-effects) model framework, and control for differences in study methods, sampling design and species-level research effort.
Research sample	All data used in this study were sourced from open-source repositories. The ecological communities data come from the PREDICTS database, a repository of 666 published studies that sampled ecological communities across land use gradients. The host-pathogen data were collated from 5 published databases or studies: the Enhanced Infectious Diseases 2 (EID2) database, Olival et al's mammal viruses database (published Nature 2017), the Global Mammal Parasite Database, Han et al's rodent reservoirs database (published PNAS 2015) and Plourde et al's reservoir hosts dataset (published PLOS One 2017), and augmented with reference to the Global Infectious Disease and Epidemiology Network (GIDEON) database. These 5 databases were standardised and combined to create a comprehensive list of host-pathogen interactions, which was then matched to the PREDICTS database to be used in our analyses. For each species in PREDICTS, we accessed species citation counts from the PubMed database, from which we derived proxy estimates of disease-related research effort. Full database descriptions are included in Methods.
Sampling strategy	Sample sizes (i.e. number of sites per land use class) were determined by ecological communities data availability within the combined PREDICTS/pathogens database. The original PREDICTS database was designed to ensure as representative sample as possible of different land use types and intensities, and the subset of data used in our analyses contains a sufficiently large number of sites per land use class to reliably detect differences (range from 369 sites for urban, to 2880 for primary).
Data collection	Ecological communities data were originally collected by the original study participants, and later collated into a single database by the PREDICTS project. Host-pathogen data were collated by the original database creators using information from surveillance data and the scientific literature. Data on disease-related research effort (used to control for species-level sampling bias) were acquired in this study by querying the PubMed online database.
Timing and spatial scale	Species occurrence/abundance data in PREDICTS were all sampled at the local (site-level) spatial scale. The dates of data collection for studies included in this analysis are between 1986 and 2013, with a median year of 2005. Full information on the scope of the PREDICTS database is included in its original data paper (cited in Supplementary Table 8).
Data exclusions	The full PREDICTS database contains species records from many studies that did not sample relevant taxonomic groups for our study focus (zoonotic disease hosts). To account for this and reduce analytical difficulties associated with zero-inflation, during data processing we excluded studies that did not sample relevant taxonomic groups: we retained any studies that sampled mammals and birds (as the major reservoir hosts of zoonoses), and for other taxa, we retained any studies that detected at least one zoonotic host in at least one site. All records of domesticated species (as defined in the EID2 database) were also excluded since these could artificially influence the results for human-modified land uses. The full data processing pipeline and rationale for these exclusions is described in Methods. Exclusion criteria were not pre-registered prior to the study commencement, but were designed and agreed prior to statistical analysis.
Reproducibility	All code and data (where not freely available online) are provided in the accompanying Figshare repository, sufficient to reproduce the results as reported. The ecological communities data are the only such large dataset available, so testing for reproducibility using an independent dataset was not possible. However, we evaluated the robustness of our main results through several sensitivity tests involving stricter subsets of the data and cross-validation, and find that the key results are consistent when zoonotic host status is more strictly defined (based on strict pathogen detection criteria), and when data are systematically excluded (either randomly or geographically). Qualitative results were also consistent when modelling three different host diversity processes (community-level, species-level, and relationship of occurrence and pathogen richness).
Randomization	Large-scale ecological data are highly non-independent as a result of underlying environmental and sampling factors that cannot be fully controlled for in field study design, therefore the PREDICTS database has a hierarchical structure with information on grouping factors in the database (multiple sites nested within studies, each of which used a standardised sampling procedure across sites). We used Bayesian mixed-effects models to account for this hierarchical structure in our statistical inference, by incorporating random intercepts accounting for study methods, spatial layout of sites within studies, and biome. We also tested the sensitivity of our main results to systematic (random and geographically-structured) downsampling of the full dataset.

Blinding	This study was based on existing databases inclusive of all associated data and metadata, and our analyses were statistical models based on this dataset. It was therefore not possible to incorporate blinding into the design of data collection and analyses.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Article

The tuatara genome reveals ancient features of amniote evolution

https://doi.org/10.1038/s41586-020-2561-9

Received: 5 December 2019

Accepted: 26 June 2020

Open access

Neil J. Gemmell[1]✉, Kim Rutherford[1], Stefan Prost[2,3], Marc Tollis[4,5], David Winter[6], J. Robert Macey[7], David L. Adelson[8], Alexander Suh[9,10], Terry Bertozzi[8,11], José H. Grau[12,13], Chris Organ[14], Paul P. Gardner[15], Matthieu Muffato[16], Mateus Patricio[16], Konstantinos Billis[16], Fergal J. Martin[16], Paul Flicek[16], Bent Petersen[17], Lin Kang[18], Pawel Michalak[18,19,20], Thomas R. Buckley[21,22], Melissa Wilson[4], Yuanyuan Cheng[23], Hilary Miller[24], Ryan K. Schott[25], Melissa D. Jordan[26], Richard D. Newcomb[26], José Ignacio Arroyo[27], Nicole Valenzuela[28], Tim A. Hore[1], Jaime Renart[29], Valentina Peona[9,10], Claire R. Peart[9,30], Vera M. Warmuth[9,30], Lu Zeng[8], R. Daniel Kortschak[8], Joy M. Raison[8], Valeria Velásquez Zapata[28], Zhiqiang Wu[28], Didac Santesmasses[31], Marco Mariotti[31], Roderic Guigó[31], Shawn M. Rupp[4], Victoria G. Twort[21,22], Nicolas Dussex[1], Helen Taylor[1], Hideaki Abe[1], Donna M. Bond[1], James M. Paterson[32], Daniel G. Mulcahy[33], Vanessa L. Gonzalez[33], Charles G. Barbieri[7], Dustin P. DeMeo[7], Stephan Pabinger[34], Tracey Van Stijn[35], Shannon Clarke[35], Oliver Ryder[36], Scott V. Edwards[37], Steven L. Salzberg[38], Lindsay Anderson[39], Nicola Nelson[39], Clive Stone[40] & Ngatiwai Trust Board*

The tuatara (*Sphenodon punctatus*)—the only living member of the reptilian order Rhynchocephalia (Sphenodontia), once widespread across Gondwana[1,2]—is an iconic species that is endemic to New Zealand[2,3]. A key link to the now-extinct stem reptiles (from which dinosaurs, modern reptiles, birds and mammals evolved), the tuatara provides key insights into the ancestral amniotes[2,4]. Here we analyse the genome of the tuatara, which—at approximately 5 Gb—is among the largest of the vertebrate genomes yet assembled. Our analyses of this genome, along with comparisons with other vertebrate genomes, reinforce the uniqueness of the tuatara. Phylogenetic analyses indicate that the tuatara lineage diverged from that of snakes and lizards around 250 million years ago. This lineage also shows moderate rates of molecular evolution, with instances of punctuated evolution. Our genome sequence analysis identifies expansions of proteins, non-protein-coding RNA families and repeat elements, the latter of which show an amalgam of reptilian and mammalian features. The sequencing of the tuatara genome provides a valuable resource for deep comparative analyses of tetrapods, as well as for tuatara biology and conservation. Our study also provides important insights into both the technical challenges and the cultural obligations that are associated with genome sequencing.

---

[1]Department of Anatomy, University of Otago, Dunedin, New Zealand. [2]LOEWE-Center for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany. [3]South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa. [4]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [5]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA. [6]School of Fundamental Sciences, Massey University, Palmerston North, New Zealand. [7]Peralta Genomics Institute, Oakland, CA, USA. [8]School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia. [9]Department of Ecology and Genetics – Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden. [10]Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden. [11]Evolutionary Biology Unit, South Australian Museum, Adelaide, South Australia, Australia. [12]Amedes Genetics, Amedes Medizinische Dienstleistungen, Berlin, Germany. [13]Museum für Naturkunde Berlin, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Berlin, Germany. [14]Department of Earth Sciences, Montana State University, Bozeman, MT, USA. [15]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [16]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. [17]Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [18]Edward Via College of Osteopathic Medicine, Blacksburg, VA, USA. [19]Center for One Health Research, Virginia–Maryland College of Veterinary Medicine, Blacksburg, VA, USA. [20]Institute of Evolution, University of Haifa, Haifa, Israel. [21]Manaaki Whenua - Landcare Research, Auckland, New Zealand. [22]School of Biological Sciences, The University of Auckland, Auckland, New Zealand. [23]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia. [24]Biomatters, Auckland, New Zealand. [25]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [26]The New Zealand Institute for Plant and Food Research, Auckland, New Zealand. [27]Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile. [28]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA. [29]Instituto de Investigaciones Biomédicas 'Alberto Sols' CSIC-UAM, Madrid, Spain. [30]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilian University of Munich, Planegg-Martinsried, Germany. [31]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain. [32]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. [33]Global Genome Initiative, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [34]Austrian Institute of Technology (AIT), Center for Health and Bioresources, Molecular Diagnostics, Vienna, Austria. [35]AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand. [36]San Diego Zoo Institute for Conservation Research, Escondido, CA, USA. [37]Department of Organismic and Evolutionary Biology and the Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA. [38]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [39]School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand. [40]Ngatiwai Trust Board, Whangarei, New Zealand. *A list of members and their affiliations appears at the end of the paper. ✉e-mail: neil.gemmell@otago.ac.nz

The tuatara is an iconic terrestrial vertebrate that is unique to New Zealand[2]. The tuatara is the only living member of the archaic reptilian order Rhynchocephalia (Sphenodontia), which last shared a common ancestor with other reptiles at about 250 million years ago (Fig. 1); this species represents an important link to the now-extinct stem reptiles from which dinosaurs, modern reptiles, birds and mammals evolved, and is thus important for our understanding of amniote evolution[2].

It is also a species of importance in other contexts. First, the tuatara is a *taonga* (special treasure) for Māori, who hold that tuatara are the guardians of special places[2]. Second, the tuatara is internationally recognized as a critically important species that is vulnerable to extinction owing to habitat loss, predation, disease, global warming and other factors[2]. Third, the tuatara displays a variety of morphological and physiological innovations that have puzzled scientists since its first description[2]. These include a unique combination of features that are shared variously with lizards, turtles and birds, which left its taxonomic position in doubt for many decades[2]. This taxonomic conundrum has largely been addressed using molecular approaches[4], but the timing of the split of the tuatara from the lineage that forms the modern squamates (lizards and snakes), the rate of evolution of tuatara and the number of species of tuatara remain contentious[2]. Finally, there are aspects of tuatara biology that are unique within, or atypical of, reptiles. These include a unique form of temperature-dependent sex determination (which sees females produced below, and males above, 22 °C), extremely low basal metabolic rates and considerable longevity[2].

To provide insights into the biology of the tuatara, we have sequenced its genome in partnership with Ngātiwai, the Māori *iwi* (tribe) who hold *kaitiakitanga* (guardianship) over the tuatara populations located on islands in the far north of New Zealand. This partnership—which, to our knowledge, is unique among the genome projects undertaken to date—had a strong practical focus on developing resources and information that will improve our understanding of the tuatara and aid in future conservation efforts. It is hoped that this work will form an exemplar for future genome initiatives that aspire to meet access and benefit-sharing obligations to Indigenous communities.

We find that the tuatara genome—as well as the animal itself—is an amalgam of ancestral and derived characteristics. Tuatara has $2n = 36$ chromosomes in both sexes, consisting of 14 pairs of macrochromosomes and 4 pairs of microchromosomes[5]. The genome size, which is estimated to be approximately 5 Gb, is among the largest of the vertebrate genomes sequenced to date; this is predominantly explained by an extraordinary diversity of repeat elements, many of which are unique to the tuatara.

Sequencing, assembly, synteny and annotation

Our tuatara genome assembly is 4.3 Gb, consisting of 16,536 scaffolds with an N50 scaffold length of 3 Mb (Extended Data Table 1, Supplementary Information 1). Genome assessment using Benchmarking Universal Single-Copy Orthologs (BUSCO)[6] indicates 86.8% of the vertebrate gene set are present and complete. Subsequent annotation identified 17,448 genes, of which 16,185 are one-to-one orthologues (Supplementary Information 2). Local gene-order conservation is high; 75% or more of tuatara genes showed conservation with birds, turtles and crocodilians. We also find that components of the genome, of 15 Mb in size and larger, are syntenic with other vertebrates; protein-coding gene order and orientation are maintained between tuatara, turtle, chicken and human, and strong co-linearity is seen between tuatara contigs and chicken chromosomes (Extended Data Figs. 1, 2).

Genomic architecture

At least 64% of the tuatara genome assembly is composed of repetitive sequences, made up of transposable elements (31%) and low-copy-number segmental duplications (33%). Although the total transposable element content is similar to other reptiles[7], the types of repeats we found appear to be more mammal-like than reptile-like. Furthermore, a number of the repeat families show evidence of recent activity and greater expansion and diversity than seen in other vertebrates (Fig. 2).

L2 elements account for most of the long interspersed elements in the tuatara genome (10% of the genome), and some may still be active (Supplementary Information 4). CR1 elements—the dominant long interspersed element in the genomes of other sauropsids[8]—are rare. CR1 elements comprise only about 4% of the tuatara genome (Fig. 2a, Supplementary Table 4.1), but some are potentially active (Supplementary Fig. 4.4). L1 elements, which are prevalent in placental mammals, account for only a tiny fraction of the tuatara genome (<1%) (Supplementary Table 4.1). However, we find that an L2 subfamily that is present in the tuatara, but is absent from other lepidosaurs, is also common in monotremes[9] (Supplementary Figs. 4.3–4.5). Collectively, these data suggest that stem-sauropsid ancestors had a repeat composition that was very different from that inferred in previous comparisons using mammals, birds and lizards[7].

Many of the short interspersed elements (SINEs) in the tuatara are derived from ancient common sequence motifs (CORE-SINEs), which are present in all amniotes[10]; however, at least 16 SINE subfamilies were recently active in the tuatara genome (Fig. 2b, Supplementary Information 5). Most of these SINEs are mammalian-wide interspersed repeats (MIRs), and the diversity of MIR subfamilies in the tuatara is the highest thus far observed in an amniote[11,12]. In the human genome, hundreds of fossil MIR elements act as chromatin and regulatory domains[13]; the very recent activity of diverse MIR subfamilies in the tuatara suggests these subfamilies may have influenced regulatory rewiring on rather recent evolutionary timescales.

We detected 24 newly identified and unique families of DNA transposon, which suggests frequent germline infiltration by DNA transposons through horizontal transfer in the tuatara[14]. At least 30 subfamilies of DNA transposon were recently active, spanning a diverse range of cut-and-paste transposons and polintons (Supplementary Figs. 5.1, 5.2). This diversity is higher than that found in other amniotes[15]. Notably, we found thousands of identical DNA transposon copies, which suggests very recent—and/or ongoing—activity. Cut-and-paste transposition probably shapes the tuatara genome, as it does in bats[15].

We identified about 7,500 full-length, long-terminal-repeat retro-elements (including endogenous retroviruses), which we classified into 12 groups (Fig. 2c, Supplementary Information 6). The general spectrum of long-terminal-repeat retroelements in the tuatara is comparable to that of other sauropsids[7,15]. We found at least 37 complete spumaretroviruses, which are among the most ancient of endogenous retroviruses[16], in the tuatara genome (Fig. 2c, Supplementary Figs. 6.1, 6.2).

The tuatara genome contains more than 8,000 elements related to non-coding RNA. Most of these elements (about 6,900) derive from recently active transposable elements, and overlap with a newly identified CR1-mobilized SINE (Fig. 2b, Supplementary Information 7). The remaining high-copy-number elements are sequences closely related to ribosomal RNAs, spliceosomal RNAs and signal-recognition particle RNAs.

Finally, a high proportion (33%) of the tuatara genome originates from low-copy-number segmental duplications; 6.7% of these duplications are of recent origin (on the basis of their high level of sequence identity (>94% identity)), which is more than seen in other vertebrates[9]. The tuatara genome is 2.4× larger than the anole genome, and this difference appears to be driven disproportionately by segmental duplications.

Overall, the repeat architecture of the tuatara is—to our knowledge—unlike anything previously reported, showing a unique amalgam of features that have previously been viewed as characteristic of either reptilian or mammalian lineages. This combination of ancient amniote

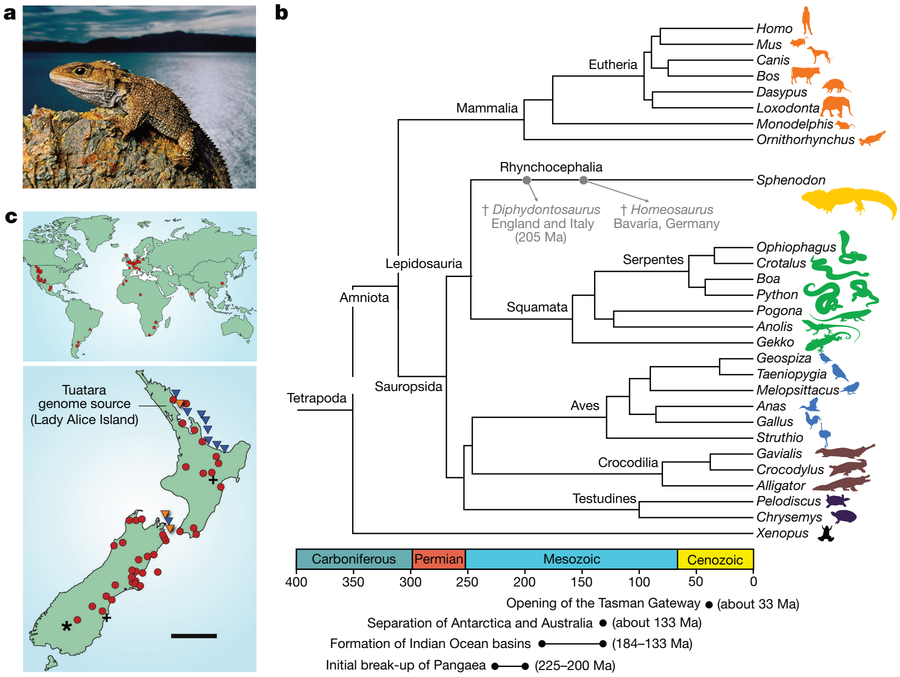


Fig. 1 | The phylogenetic significance and distribution of the tuatara. a, The tuatara, (*S. punctatus*) is the sole survivor of the order Rhynchocephalia. **b**, **c**, The rhynchocephalians appear to have originated in the early Mesozoic period (about 250–240 million years ago (Ma)) and were common, speciose and globally distributed for much of that era. The geographical range of the rhynchocephalians progressively contracted after the Early Jurassic epoch (about 200–175 Ma); the most recent fossil record outside of New Zealand is from Argentina in the Late Cretaceous epoch (about 70 Ma). **c**, The last bastions of the rhynchocephalians are 32 islands off the coast of New Zealand, which have recently been augmented by the establishment of about 10 new island or mainland sanctuary populations using translocations. The current global population is estimated to be around 100,000 individuals. Rhynchocephalian and tuatara fossil localities are redrawn and adapted from ref.[1] with permission, and incorporate data from ref.[2]. In the global distribution map (**c**, top); triangle = Triassic; square = Jurassic; circle = Cretaceous; and diamond = Palaeocene. In the map of the New Zealand distribution (**c**, bottom); asterisk = Miocene; cross = Pleistocene; circle = Holocene; blue triangle = extant population; and orange triangle = population investigated in this study. Scale bar, 200 km. Photograph credit, F. Lanting.

features—as well as a dynamic and diverse repertoire of lineage-specific transposable elements—strongly reflects the phylogenetic position of this evolutionary relic.

Our low-coverage bisulfite-sequencing analysis found approximately 81% of CpG sites are methylated in tuatara (Fig. 3a)—the highest reported percentage of methylation for an amniote. This pattern differs from that observed in mouse, human (about 70%) and chicken (about 50%), and is more similar to that of *Xenopus* (82%) and zebrafish (78%). One possible explanation for this high level of DNA methylation is the large number of repetitive elements found in tuatara, many of which appear recently active and might be regulated via DNA methylation.

The low normalized CpG content of the tuatara suggests its genome has endured substantial historic methylation[17]. The tuatara has a significantly bimodal distribution of normalized CpG (Extended Data Fig. 3) in all of the genomic regions we examined, a similarity it shares with other reptiles that have temperature-dependent sex determination[17]. The low normalized CpG count of the tuatara in non-promoter regions may result from methylation silencing of repeat elements, and the bimodality of normalized CpG promoters suggests dual transcriptional regulation (Extended Data Fig. 3, Supplementary Information 8).

The mitochondrial genome in the tuatara reference animal is 18,078 bp in size, containing 13 protein-coding, 2 ribosomal RNA and 22 transfer (t)RNA genes, a gene content typical among animals (Extended Data Fig. 4). This contradicts previous reports[18] that the tuatara mitochondrial genome lacks three genes: *ND5*, *tRNA*^{Thr} and *tRNA*^{His}. These genes are found—with an additional copy of *tRNA*$^{Leu(CUN)}$ and an additional non-coding block (which we refer to as NC2)—in a single segment of the mitochondrial genome. Three non-coding areas (NC1, NC2 and NC3) with control-region (heavy-strand replication origin) features, and two copies of *tRNA*$^{Leu(CUN)}$ adjacent to NC1 and NC2, possess identical or near-identical sequences that are unique to the tuatara mitochondrial genome. These three non-coding regions may be a result of concerted evolution.

Genomic innovations

As befits the taxonomic distinctiveness of the tuatara, we find that its genome displays multiple innovations in genes that are associated with immunity, odour reception, thermal regulation and selenium metabolism.

Genes of the major histocompatibility complex (MHC) have an important role in disease resistance, mate choice and kin recognition, and are among the most polymorphic genes in the vertebrate genome. Our annotation of MHC regions in the tuatara, and comparisons of the gene organization with that of six other species, identified 56 MHC genes (Extended Data Fig. 5, Supplementary Information 9).

Of the six comparison species, the genomic organization of tuatara MHC genes is most similar to that of the green anole, which we interpret as typical for Lepidosauria. Tuatara and other reptiles show a gene content and complexity more similar to the MHC regions of amphibians and mammals than to the highly reduced MHC of birds. Although the majority of genes annotated in the tuatara MHC are well-conserved as one-to-one orthologues, we observed extensive genomic rearrangements among these distant lineages.

The tuatara is a highly visual predator that is able to capture prey under conditions of extremely low light[2]. Despite the nocturnal visual adaptation of the tuatara, it shows strong morphological evidence of an ancestrally diurnal visual system[19]. We identified all five of the

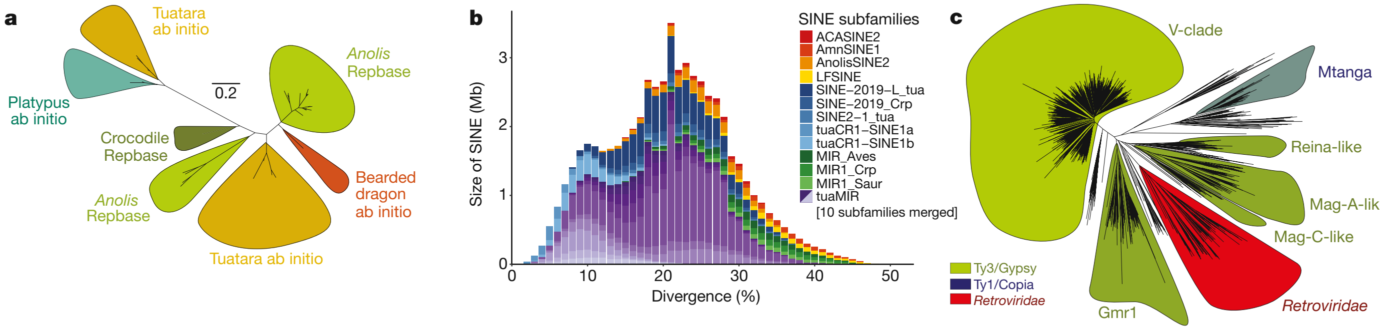


Fig. 2 | Analysis of the repeat landscape in the tuatara genome identifies unique repeat families, evidence of recent activity and a greater expansion and diversity of repeats than any other amniote. a, A phylogenetic analysis on the basis of the reverse transcriptase domain of L2 repeats identifies two L2 subfamilies; one typical of other lepidosaurs and one that is similar to platypus L2. This phylogeny is based on L2 elements >1.5-kb long with a reverse transcriptase domain of >200 amino acids. **b**, Landscape plot of SINE retrotransposons suggests the tuatara genome is dominated by MIR sequences that are most typically associated with mammals; the tuatara genome is now the amniote genome in which the greatest MIR diversity has been observed. Only SINE subfamilies that occupy more than 1,000 bp are shown. Definitions of the abbreviations of the SINE subfamilies follow: ACASINE2, *Anolis carolinesis* SINE family; AmnSINE1, Amniota SINE1; AnolisSINE2, *A. carolinesis* SINE2 family, LFSINE, lobe-finned fishes SINE; SINE−2019−L_tua, tuatara SINE; SINE-2019_Crp, *Crocodylus porosus* SINE; SINE2-1_tua, tuatara SINE2; tuaCR1-SINE1a and b, tuatara CR1-mobilized SINEs; MIR_Aves, avian MIR sequence; MIR1_Crp, *C. porosus* MIR sequence; MIR1_Saur, Sauropsida MIR sequence; tuaMIR, tuatara MIRs. **c**, The tuatara genome contains about 7,500 full-length, long-terminal-repeat retro-elements, including nearly 450 endogenous retroviruses that span the five major retroviral clades. A Ty1/Copia element (Mtanga-like) is especially abundant, but Bel-Pao long-terminal-repeat retro-elements are absent. At least 37 complete spumaretroviruses are present in the tuatara genome.

vertebrate visual opsin genes in the tuatara genome (Supplementary Information 10).

Our comparative analysis revealed one of the lowest rates of visual-gene loss known for any amniote, which contrasts sharply with the high rates of gene loss observed in ancestrally nocturnal lineages (Extended Data Fig. 6). Visual genes involved in phototransduction showed strong negative selection and no evidence for the long-term shifts in selective pressures that have been observed in other groups with evolutionarily modified photoreceptors[20]. The retention of five visual opsins and the conserved nature of the visual system also suggests tuatara possess robust colour vision, potentially at low light levels. This broad visual repertoire may be explained by the dichotomy in tuatara life history: juvenile tuatara often take up a diurnal and arboreal lifestyle to avoid the terrestrial, nocturnal adults that may predate them[2]. Collectively, these results suggest a unique path to nocturnal adaptation in tuatara from a diurnal ancestor.

Odorant receptors are expressed in the dendritic membranes of olfactory receptor neurons and enable the detection of odours. Species that depend strongly on their sense of smell to interact with their environment, find prey, identify kin and avoid predators may be expected to have a large number of odorant receptors. The tuatara genome contains 472 predicted odorant receptors, of which 341 sequences appear intact (Supplementary Information 11). The remainder lack the initial start codon, have frameshifts or are presumed to be pseudogenes. Many odorant receptors were found as tandem arrays, with up to 26 genes found on a single scaffold.

The number and diversity of odorant receptor genes varies greatly in Sauropsida: birds have 182–688 such genes, the green anole lizard has 156 genes, and crocodilians and testudines have 1,000–2,000 genes[21]. The tuatara has a number of odorant receptors similar to that of birds, but contains a high percentage of intact odorant receptor genes (85%) relative to published odorant receptor sets from the genomes of other sauropsids. This may reflect a strong reliance on olfaction by tuatara, and therefore pressure to maintain a substantial repertoire of odorant receptors (Extended Data Fig. 7). There is some evidence that olfaction has a role in identifying prey[2], as well as suggestions that cloacal secretions may act as chemical signals.

The tuatara is a behavioural thermoregulator, and is notable for having the lowest optimal body temperature of any reptile (16–21 °C). Genes that encode transient receptor potential ion channels (TRP genes) have an important role in thermoregulation, as these channels participate in thermosensation and cardiovascular physiology[22]; this led us to hypothesize that TRP genes may be linked to the thermal tolerance of the tuatara. Our comparative genomic analysis of TRP genes in the tuatara genome identified 37 TRP-like sequences, spanning all 7 known subfamilies of TRP genes (Extended Data Fig. 8, Supplementary Information 12)— an unusually large repertoire of TRP genes.

Among this suite of genes, we identified thermosensitive and non-thermosensitive TRP genes that appear to result from gene duplication, and have been differentially retained in the tuatara. For example, the tuatara is unusual in possessing an additional copy of a thermosensitive TRPV-like gene (*TRPV1/2/4*, sister to the genes *TRPV1*, *TRPV2* and *TRPV4*) that has classically been linked to the detection of moderate-to-extreme heat[22]—a feature it shares with turtles. A strong signature of positive selection among heat-sensitive TRP genes (*TRPA1*, *TRPM* and *TRPV*) was also observed.

In general, these results show a high rate of differential retention and positive selection in genes for which a function in heat sensation is well-established[22]. It therefore seems probable that the genomic changes in TRP genes are associated with the evolution of thermoregulation in tuatara.

Barring tortoises, tuatara are the longest lived of the reptiles—probably exceeding 100 years of age[2]. This enhanced lifespan may be linked to genes that afford protection against reactive oxygen species. One class of gene products that affords such protection is the selenoproteins. The human genome encodes 25 selenoproteins, the roles of which include antioxidation, redox regulation, thyroid hormone synthesis and calcium signal transduction, among others[23].

We identified 26 genes that encode selenoproteins in the tuatara genome, as well as 4 selenocysteine-specific tRNA genes; all of these appear to be functional (Supplementary Information 13). Although further work is needed, the additional selenoprotein gene (relative to the human genome) and the selenocysteine-specific tRNA genes may be linked to the longevity of tuatara or might have arisen as a response to the low levels of selenium and other trace elements in the terrestrial systems of New Zealand.

Tuatara has a unique mode of temperature-dependent sex determination, in which higher temperatures during egg incubation result in males[2]. We found orthologues for many genes that are known to act antagonistically in masculinizing (for example, *SF1* and *SOX9*) and feminizing (for

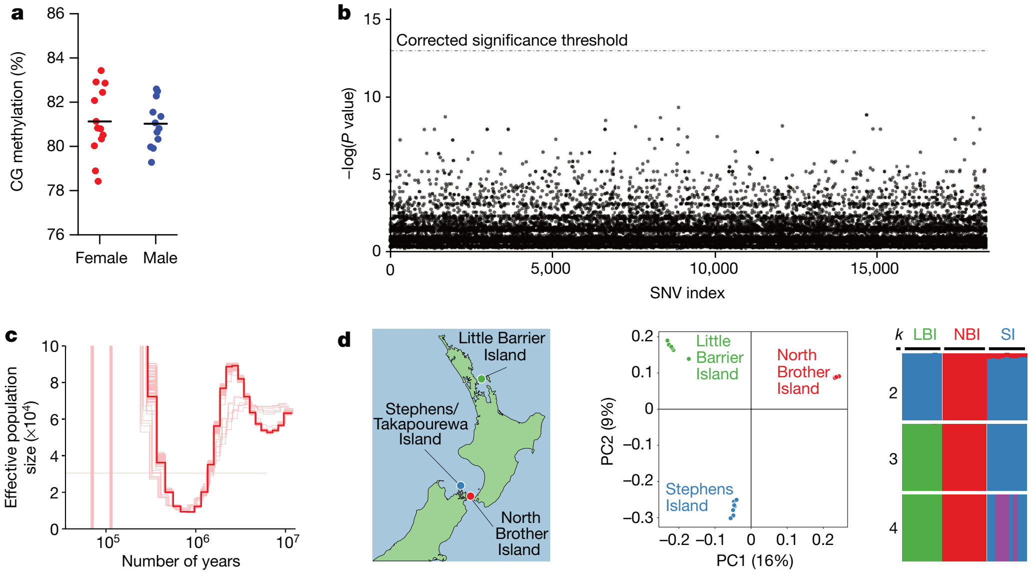


Fig. 3 | Analysis of sex differences, demographic history and population structure. a, Methylation levels in the tuatara genome are high (mean 81%), but show no significant differences among the sexes (female $n = 13$, mean = 81.13, s.d. = 1.55; male $n = 12$; mean = 81.02, s.d.–1.07). The black horizontal line represents the mean in each dataset. **b**, No single-nucleotide variant (SNV) is significantly differentiated with respect to sex in the tuatara genome. Each point represents a P value from a test of sexual differentiation for a single SNV. The dashed line represents the threshold for statistical significance after accounting for multiple testing ($n = 28$; 13 males and 15 females). P values calculated using Fisher's exact test, two-tailed test and corrected for multiple testing using the Bonferroni method. **c**, Pairwise sequential Markovian coalescent plot of the demographic history of tuatara using a mutation rate of 1.4×10^{-8} substitutions per site per generation and a generation time of 30 years. **d**, We examined the three known axes of genetic diversity in tuatara: northern New Zealand (Little Barrier Island (LBI) ($n = 9$)) and two islands in the Cook Strait (Stephens Island (SI) ($n = 9$) and North Brother Island (NBI) ($n = 10$)), using genotype-by-sequencing methods. Principal component (PC) analysis and structure plots demonstrate substantial structure among tuatara populations, and strongly support previous suggestions that the tuatara on the North Brother Island are genetically distinct and warrant separate management.

example, *RSPO1* and *WNT4*) gene networks to promote testicular or ovarian development, respectively[24]. We also found orthologues of several genes that have recently been implicated in temperature-dependent sex determination, including *CIRBP*[24] (Supplementary Information 17, Supplementary Table 17.2). Tuatara possess no obviously differentiable sex chromosomes[5], and we found no significant sex-specific differences in global CG methylation (Fig. 3a) and no sex-specific single-nucleotide variants between male and female tuatara (Fig. 3b). On a gene-by-gene basis, sex-specific differences in methylation and gene expression patterns probably exist, but this remains to be investigated.

Phylogeny and evolutionary rates

Our phylogenomic analyses, which incorporated both whole-genome alignments and clusters of single-copy orthologues (Supplementary Information 14, 15) recapitulated many patterns that have been observed in the fossil record and corroborated during the genomic era (Fig. 1). After their appearance about 312 million years ago[25], amniote vertebrates diversified into two groups: the synapsids (which include all mammals) and the sauropsids (which include all reptiles and birds). We obtained full phylogenomic support for a monophyletic Lepidosauria, marked by the divergence of the tuatara lineage from all squamates (lizards and snakes) during the early part of the Triassic period at about 250 million years ago, as estimated using a penalized likelihood method (Fig. 1, Supplementary Information 14–16).

The rate of molecular evolution in the tuatara has previously been suggested to be paradoxically high, in contrast to the apparently slow rate of morphological evolution[26,27]. However, we find that the actual divergence in terms of DNA substitutions per site per million years at fourfold degenerate sites is relatively low, particularly with respect to lizards and snakes; this makes the tuatara the slowest-evolving lepidosaur yet analysed (Extended Data Fig. 9a, b). We also find that in general amniote evolution can be described by a model of punctuated evolution, in which the amount of genomic change is related to the degree of species diversification within clades[28,29]. The tuatara falls well below this trend, accumulating substitutions at a rate expected given the lack of rhynchocephalian diversity (Extended Data Fig. 9c, Supplementary Information 16). This suggests that rates of phenotypic and molecular evolution were not decoupled throughout the evolution of amniotes[30].

Patterns of selection

In two sets of analyses, we find that most genes exhibit a pattern of molecular evolution that suggests that the tuatara branch evolves at a different rate than the rest of the tree (Supplementary Information 17, Supplementary Table 4). Approximately 659 of the 4,284 orthologues we tested had significantly different ω values (ratios of non-synonymous to synonymous substitutions, dN/dS) on the tuatara branch relative to the birds and other reptiles we tested (Supplementary Information 17). Although none of these orthologues had ω values suggestive of strong positive selection (that is, >1), the results do indicate that shifts in patterns of selection are affecting many genes and functional categories of genes across the tuatara genome, including genes involved in RNA regulation, metabolic pathways, general metabolism and sex determination.

Population genomics

Once widespread across the supercontinent of Gondwana, Rhynchocephalia is now represented by a single species (the tuatara) found on a few islands offshore of New Zealand (Fig. 1c). Historically, tuatara declined in range and numbers because of introduced pests and habitat loss[2]. They remain imperilled owing to their highly restricted distribution, threats imposed by disease and changes in sex ratios induced by climate change that could markedly affect their survival[31]. Previous work has found that populations in northern New Zealand are genetically distinct from those in the Cook Strait, and that the population on North Brother Island in the Cook Strait might be a distinct species[3]. Although subsequent studies have not supported species status for the population on North Brother Island[32], it is managed as a separate conservation unit.

We used the tuatara reference genome to perform ancestral demographic and population genomic analyses of this species. First, we investigated genome-wide signals for demographic change using a pairwise sequentially Markovian coalescent method (Supplementary Information 18). Our reconstructed demography (Fig. 3c) reveals an increase in effective population size (N_e) that is detectable around 10 million years ago, a marked decrease in N_e about 1–3 million years ago and a rapid increase in N_e between 500 thousand years ago and 1 million years ago. These events correlate well with the known geological history of New Zealand[33], and may reflect an increase in available landmass subsequent to Oligocene drowning, a period of considerable climatic cooling that probably reduced tuatara habitat and the formation of land bridges that facilitated population expansion.

Our population genomic analyses examined the major axes of genetic diversity in tuatara[32,34], and revealed substantial genetic structure (Fig. 3d, Supplementary Information 19). Our genome-wide estimate of the fixation index (F_{ST}) is 0.45, and more than two-thirds of variable sites have an allele that is restricted to a single island. All populations have relatively low genetic diversity (nucleotide diversity ranges from 8×10^{-4} for North Brother Island to 1.1×10^{-3} for Little Barrier Island). The low within-population diversity and marked population structure we observe in the tuatara suggests that the modern island populations were isolated from each other sometime during the Last Glacial Maximum at about 18 thousand years ago.

Our results also support the distinctiveness of the North Brother Island tuatara, which has variously been described as *S. punctatus* or *Sphenodon guntheri*[3,32]. This population is highly inbred and shows evidence of a severe bottleneck, which most probably reflects a founder event around the time of the last glaciation[34]. It is not clear whether the distinctiveness we observe is due to changes in allele frequency brought about by this bottleneck, or is reflective of a deeper split in the population history of tuatara. Regardless, this population is an important source of genetic diversity in tuatara, possessing 8,480 private alleles. Although we support synonymization of *S. punctatus* and *S. guntheri*[32], the ongoing conservation of the North Brother Island population as an independent unit is recommended.

A cultural dimension

The tuatara is a *taonga* for many Māori—notably Ngātiwai and Ngāti Koata who are the *kaitiaki* (guardians) of tuatara. We worked in partnership with Ngātiwai *iwi* to increase knowledge and understanding of tuatara, and aid in the conservation of this species in the long term. Ngātiwai were involved in all decision-making regarding the use of the genome data by potential collaborators; for each new project we proposed, we discussed the benefits that might accrue from this work and how these could be shared. The need to engage with—and protect the rights of—Indigenous communities in such a transparent way has seldom been considered in the genome projects published to date, but is a mandated consideration under the Nagoya Protocol (https://www.cbd.int/abs/). Our partnership is a step towards an inclusive model of genomic science, which we hope others will adopt and improve upon. Although each partnership is unique, we provide a template agreement (Supplementary Information 20) that we hope will be useful to others.

Discussion

The tuatara has a genomic architecture unlike anything previously reported, with an amalgam of features that have previously been viewed as characteristic of either mammals or reptiles. Notable among these features are unusually high levels of repetitive sequences that have traditionally been considered mammalian, many of which appear to have been recently active, and—to our knowledge—the highest level of genome methylation thus far reported. We also found a mitochondrial-genome gene content at odds with previously published reports that omitted the *ND5* gene[18]; this gene is present, nested within a repeat-rich region of the mitochondrial DNA.

Our phylogenetic studies provide insights into the timing and speed of amniote evolution, including evidence of punctuated genome evolution across this phylogeny. We also find that, in contrast to previous suggestions that the evolutionary rate for tuatara is exceptionally fast[26], it is the slowest-evolving lepidosaur yet analysed.

Our investigations of genomic innovations identified genetic candidates that may explain the ultra-low active body temperature, longevity and apparent resistance to infectious disease in tuatara. Further functional exploration will refine our understanding of these unusual facets of tuatara biology, and the tuatara genome itself will enable many future studies to explore the evolution of complex systems across the vertebrates in a more complete way than has previously been possible.

Our population genomic work reveals considerable genetic differences among tuatara populations, and supports the distinctiveness of the North Brother Island tuatara.

Finally, this genome will greatly aid in future work on population differentiation, inbreeding and local adaptation in this global icon, the last remaining species of the once globally dominant reptilian order Rhynchocephalia.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2561-9.

1. Jones, M. E., Tennyson, A. J., Worthy, J. P., Evans, S. E. & Worthy, T. H. A sphenodontine (Rhynchocephalia) from the Miocene of New Zealand and palaeobiogeography of the tuatara (*Sphenodon*). *Proc. R. Soc. Lond. B* **276**, 1385–1390 (2009).
2. Cree, A. *Tuatara: Biology and Conservation of a Venerable Survivor* (Canterbury Univ. Press, 2014).
3. Daugherty, C. H., Cree, A., Hay, J. M. & Thompson, M. B. Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature* **347**, 177–179 (1990).
4. Jones, M. E. H. et al. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **13**, 208 (2013).
5. O'Meally, D., Miller, H., Patel, H. R., Graves, J. A. M. & Ezaz, T. The first cytogenetic map of the tuatara, *Sphenodon punctatus*. *Cytogenet. Genome Res.* **127**, 213–223 (2009).
6. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
7. Pasquesi, G. I. M. et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* **9**, 2774 (2018).
8. Suh, A. et al. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol. Evol.* **7**, 205–217 (2015).
9. Zeng, L., Kortschak, R. D., Raison, J. M., Bertozzi, T. & Adelson, D. L. Superior ab initio identification, annotation and characterisation of TEs and segmental duplications from genome assemblies. *PLoS ONE* **13**, e0193588 (2018).
10. Gilbert, N. & Labuda, D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl Acad. Sci. USA* **96**, 2869–2874 (1999).
11. Mikkelsen, T. S. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
12. Warren, W. C. et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
13. Wang, J. et al. MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl Acad. Sci. USA* **112**, E4428–E4437 (2015).
14. Schaack, S., Gilbert, C. & Feschotte, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**, 537–546 (2010).
15. Sotero-Caio, C. G., Platt, R. N. II, Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
16. Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat. Commun.* **8**, 13954 (2017).
17. Radhakrishnan, S., Literman, R., Mizoguchi, B. & Valenzuela, N. MeDIP-seq and nCpG analyses illuminate sexually dimorphic methylation of gonadal development genes with high historic methylation in turtle hatchlings with temperature-dependent sex determination. *Epigenetics Chromatin* **10**, 28 (2017).

18. Rest, J. S. et al. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol. Phylogenet. Evol*. **29**, 289–297 (2003).
19. Meyer-Rochow, V. B., Wohlfahrt, S. & Ahnelt, P. K. Photoreceptor cell types in the retina of the tuatara (*Sphenodon punctatus*) have cone characteristics. *Micron* **36**, 423–428 (2005).
20. Schott, R. K., Bhattacharyya, N. & Chang, B. S. W. Evolutionary signatures of photoreceptor transmutation in geckos reveal potential adaptation and convergence with snakes. *Evolution* **73**, 1958–1971 (2019).
21. Vandewege, M. W. et al. Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biol. Evol*. **8**, 470–480 (2016).
22. Nilius, B. & Owsianik, G. The transient receptor potential family of ion channels. *Genome Biol*. **12**, 218 (2011).
23. Labunskyy, V. M., Hatfield, D. L. & Gladyshev, V. N. Selenoproteins: molecular pathways and physiological roles. *Physiol. Rev*. **94**, 739–777 (2014).
24. Capel, B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet*. **18**, 675–689 (2017).
25. Blair Hedges, S. & Kumar, S. *The Timetree of Life* (Oxford Univ. Press, 2009).
26. Hay, J. M., Subramanian, S., Millar, C. D., Mohandesan, E. & Lambert, D. M. Rapid molecular evolution in a living fossil. *Trends Genet*. **24**, 106–109 (2008).
27. Miller, H. C., Moore, J. A., Allendorf, F. W. & Daugherty, C. H. The evolutionary rate of tuatara revisited. *Trends Genet*. **25**, 13–15, author reply 16–18 (2009).
28. Landis, M. J. & Schraiber, J. G. Pulsed evolution shaped modern vertebrate body sizes. *Proc. Natl Acad. Sci. USA* **114**, 13224–13229 (2017).
29. Webster, A. J., Payne, R. J. H. & Pagel, M. Molecular phylogenies link rates of evolution and speciation. *Science* **301**, 478 (2003).
30. Subramanian, S., Hay, J. M., Mohandesan, E., Millar, C. D. & Lambert, D. M. Molecular and morphological evolution in tuatara are decoupled. *Trends Genet*. **25**, 16–18 (2009).
31. Mitchell, N. J., Kearney, M. R., Nelson, N. J. & Porter, W. P. Predicting the fate of a living fossil: how will global warming affect sex determination and hatching phenology in tuatara? *Proc. R. Soc. Lond. B* **275**, 2185–2193 (2008).
32. Hay, J. M., Sarre, S. D., Lambert, D. M., Allendorf, F. W. & Daugherty, C. H. Genetic diversity and taxonomy: a reassessment of species designation in tuatara (*Sphenodon*: Reptilia). *Conserv. Genet*. **11**, 1063–1081 (2010).
33. Cooper, A. & Cooper, R. A. The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis. *Proc. R. Soc. Lond. B* **261**, 293–302 (1995).
34. MacAvoy, E. S. et al. Genetic variation in island populations of tuatara (*Sphenodon* spp) inferred from microsatellite markers. *Conserv. Genet*. **8**, 305–318 (2007).

Ngatiwai Trust Board

Clive Stone[40], Jim Smillie[40] & Haydn Edmonds[40]

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

A full description of the methods can be found in the Supplementary Information.

Sampling and sequencing

A blood sample was obtained from a large male tuatara from Lady Alice Island (35° 53′ 24.4″ S 174° 43′ 38.2″ E) (New Zealand), with appropriate ethical permissions and *iwi* consultation and support (Supplementary Information 20). Total genomic DNA and RNA were extracted and sequenced using the Illumina HiSeq 2000 and MiSeq sequencing platforms (Illumina) supported by New Zealand Genomics (Supplementary Information 1).

Genome, transcriptome and epigenome

Raw reads were de novo-assembled using Allpaths-LG (version 49856). With a total input data of 5,741,034,516 reads for the paired-end libraries and 2,320,886,248 reads of the mate-pair libraries, our optimal assembly used 85% of the fragment libraries and 100% of the jumping libraries (Supplementary Information 1.4). We further scaffolded the assembly using Chicago libraries and HiRise (Supplementary Information 1.3).

We assembled a de novo transcriptome as a reference for read-mapping using total RNA derived from the blood of our reference male tuatara, and a collection of transcriptomic data previously collected from early-stage embryos[35]. In total, we had 131,580,633 new 100-bp read pairs and 60,637,100 previous 50-bp read pairs. These were assembled using Trinity v.2.2.0 (Supplementary Information 1.4).

Low-coverage bisulfite sequencing was undertaken using a modified post-bisulfite adaptor tagging method to explore global patterns of methylation in the genome for 12 male and 13 female tuatara (Fig. 3d, Supplementary Information 1.5).

Repeat and gene annotation

We used a combination of ab initio repeat identification in CARP/RepeatModeler/LTRharvest, manual curation of specific newly identified repeats, and homology to repeat databases to investigate the repeat content of the tuatara genome (Supplementary Information 1.6). From these three complementary repeat identification approaches, the CARP results were in-depth-annotated for long interspersed elements and segmental duplications (Supplementary Information 4), the RepeatModeler results were in-depth-annotated for SINEs and DNA transposons (Supplementary Information 5), and the LTRharvest results were in-depth-annotated for long-terminal-repeat retrotransposons (Supplementary Information 6).

For the gene annotation, we used RepeatMasker (v.4.0.3) along with our partially curated RepeatModeler library plus the Repbase sauropsid repeat database to mask transposable elements in the genome sequence before the gene annotation. We did not mask simple repeats at this point to allow for more efficient mapping during the homology-based step in the annotation process. Simple repeats were later soft-masked and protein-coding genes predicted using MAKER2. We used anole lizard (*A. carolinensis*, version AnoCar2.0), python (*Python bivittatus*, version bivittatus-5.0.2) and RefSeq (www.ncbi.nlm.nih.gov/refseq) as protein homology evidence, which we integrated with ab initio gene prediction methods including BLASTX, SNAP and Augustus. Non-coding RNAs were annotated using Rfam covariance models (v.13.0) (Supplementary Information 7).

Orthologue calling

We performed a phylogenetic analysis to infer orthology relationships between the tuatara and 25 other species using the Ensembl GeneTree method (Supplementary Information Tables 2.1, 2.2). Multiple-sequence alignments, phylogenetic trees and homology relationships were extracted in various formats (https://zenodo.org/record/2542571). We also calculated the gene order conservation score, which uses local synteny information around a pair of orthologous genes to compute how much the gene order is conserved. For each of these species, we chose the paralogue with the best gene order conservation score and sequence similarity, which resulted in a total set of 3,168 clusters of orthologues (Supplementary Information 2, Table 2.3).

Gene tree reconstructions and substitution rate estimation

We constructed phylogenies using only fourfold-degenerate-site data derived from whole-genome alignments for 27 tetrapods, analysed as a single partition in RAxML v.8.2.3. Using the topology and branch lengths obtained from the best maximum likelihood phylogeny, we estimated absolute rates of molecular evolution in terms of substitution per site per million years and estimated the divergence times of amniotes via the semiparametric penalized likelihood method in r8s v.1.8 (Supplementary Information 14.5).

We also generated gene trees on the basis of 245 single-copy orthologues found between all species using a maximum-likelihood-based multi-gene approach (Supplementary Information 15). Sequences were aligned using the codon-based aligner PRANK. The FASTA format alignments were then converted to PHYLIP using the catfasta2phyml.pl script (https://github.com/nylander/catfasta2phyml). Next, we used the individual exon PHYLIP files for gene tree reconstruction using RaxML using a GTR + G model. Subsequently, we binned all gene trees to reconstruct a species tree and carried out bootstrapping using Astral (Supplementary Information 15, Supplementary Fig. 15.1).

Divergence times and tests of punctuated evolution

We inferred time-calibrated phylogenies with BEAST v.2.4.8 using the CIPRES Science Gateway to explore divergence times (Supplementary Information 16.1). We then used Bayesian phylogenetic generalized least squares to regress the total phylogenetic path length (of fourfold-degenerate sites) on the net number of speciation events (nodes in a phylogenetic tree) as a test for punctuated evolution (Supplementary Information 16.2).

Analysis of genomic innovations

We explored the genomic organization and sequence evolution of genes associated with immunity, vision, smell, thermoregulation, longevity and sex determination (Supplementary Information 8–13). Tests of selection were undertaken across multiple genes, including those linked to metabolism, vision and sex determination using multispecies alignments and PAML (Supplementary Information 17).

Population genomics

Demographic history was inferred from the diploid sequence of our tuatara genome using a pairwise sequential Markovian coalescent method (Supplementary Information 18). We also sampled 10 tuatara from each of three populations that span the main axes of genetic diversity in tuatara (Supplementary Information 19, Table 19.1), and used a modified genotype-by-sequencing approach to obtain the SNVs that we used for population genomic analysis, investigations of loci associated with sexual phenotype and estimates of genetic load (Supplementary Information 19).

Permits and ethics

This project was undertaken in partnership with Ngatiwai and in consultation with other *iwi* who are *kaitiaki* of tuatara (Supplementary Information 20). Samples were collected under Victoria University of Wellington Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347 and held and used under permits 45462-DOA and 32037-RES 32037-RES issued by the New Zealand Department of Conservation.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The Tuatara Genome Consortium Project whole-genome shotgun and genome assembly are registered under the umbrella BioProjects PRJNA418887 and PRJNA445603, which are associated with BioSamples SAMN08038466 and SAMN08793959. Transcriptome read data are submitted under SRR7084910 (whole blood), together with previous data (SRR485948). The transcriptome assembly is submitted to GenBank with ID GGNQ00000000.1. Illumina short-read, Oxford Nanopore and PacBio long-read sequences are in the Sequence Read Archive accessions associated with PRJNA445603. The genome assembly (GCA_003113815.1) described in this paper is version QEPC00000000.1 and consists of sequences QEPC01000001–QEPC01016536. Maker gene predictions are available from Zenodo at https://doi.org/10.5281/zenodo.1489353. The repeat library database developed for tuatara is available from Zenodo at https://doi.org/10.5281/zenodo.2585367.

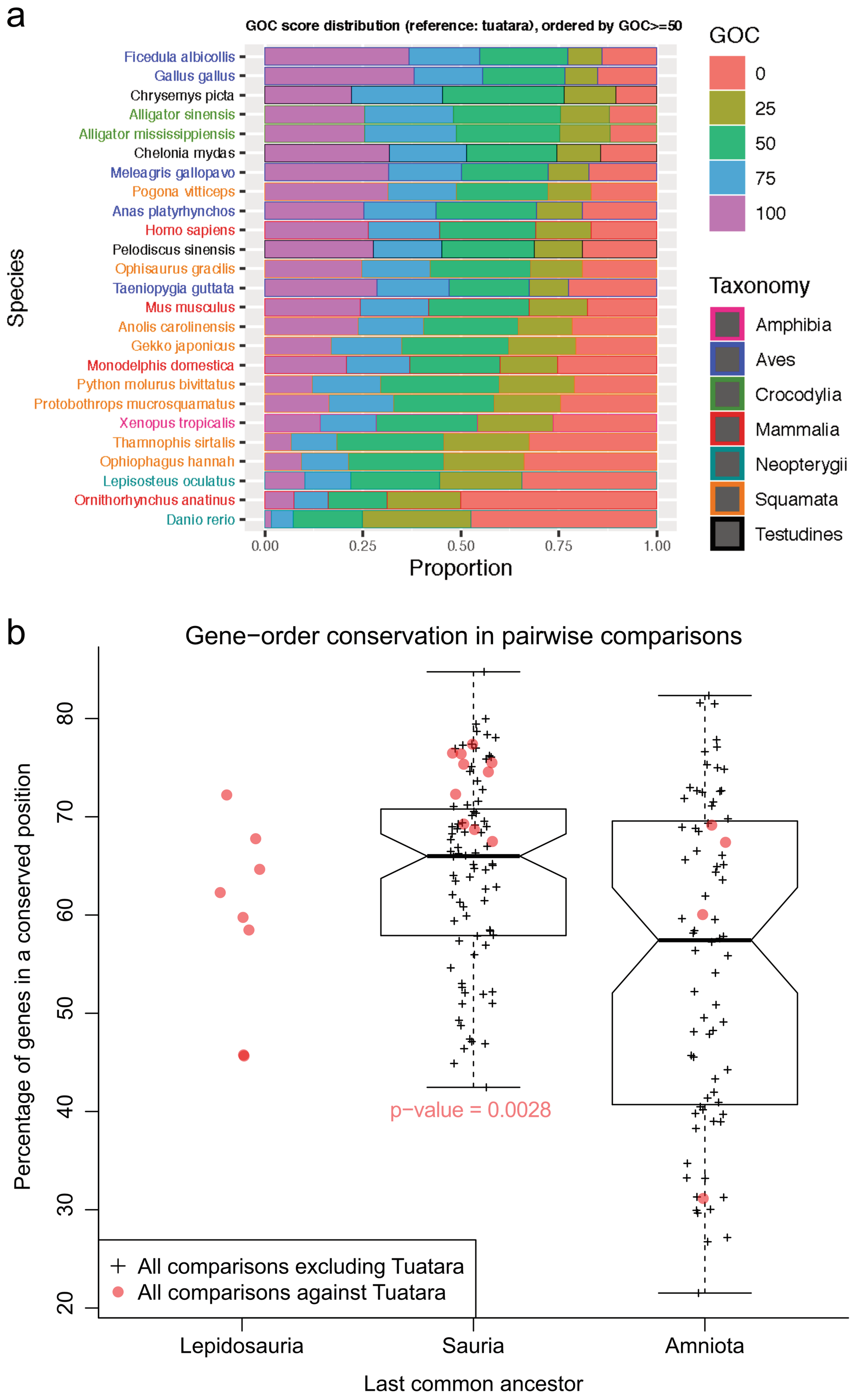
35. Miller, H. C., Biggs, P. J., Voelckel, C. & Nelson, N. J. De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (*Sphenodon punctatus*). *BMC Genomics* **13**, 439 (2012).
36. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron*. **18.1**, 18.1.1FC (2015).
37. Head, J. J. Fossil calibration dates for molecular phylogenetic analysis of snakes 1: Serpentes, Alethinophidia, Boidae, Pythonidae. *Palaeontol. Electron*. **18.1**, 18.1.6.FC (2015).

Acknowledgements N.J.G. acknowledges the support of Ngatiwai *iwi*, Allan Wilson Centre, University of Otago, New Zealand Department of Conservation, New Zealand Genomics and Illumina. J.I.A. was supported by CONICYT National Doctoral Scholarship No. 21130515. M.W. was supported by NIH grant R35 GM124827. Ensembl annotation was supported by the Wellcome Trust (WT108749/Z/15/Z) and the European Molecular Biology Laboratory. We thank Ngāti Koata, Te Ātiawa o Te Waka-a-Māui, and Ngāti Manuhiri *iwi* for granting permission to reuse tuatara samples obtained from Stephens Island (Takapourewa), North Brother Island and Little Barrier Island (Hauturu), respectively; all of the people involved in obtaining and curating the samples held in the Victoria University of Wellington tuatara collection; A. Zimin, D. Puiu, G. Marcais, J. Yorke and R. Crowhurst for help with and discussions about genome assembly; I. Fiddes, J. Armstrong and B. Paten for help with comparative genome alignments and annotation; the National eScience Infrastructure (NeSI) and Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for computational support; R. McPhee for help with figures; and T. Braisher for manuscript coordination and editing.

Author contributions N.J.G. designed the original concept and scientific objectives and oversaw the project and analyses. N.J.G., L.A., N.N., H.T., O.R., S.V.E., C.S. contributed samples or assisted in sample preparation and permitting. N.J.G., K.M.R., S.P., M.T., D.W., J.R.M., D.L.A., A.S., T.B., J.H.G., C.O., P.P.G., M.M. M.P., K.B., F.J.M., P.F., B.P., L.K., P.M., T.R.B., M.W., Y.C., H.M., R.K.S., M.D.J., R.D.N., J.I.A., N.V., T.A.H., J.R., V.P., C.R.P., V.M.W., L.Z., R.D.K., J.M.R., V.V.Z., Z.W., D.S., M. Mariotti, R.G., S.M.R., V.G.T., N.D., H.A., D.M.B, J.M.P., D.G.M., V.L.G., C.G.B., D.P.D., S. Pabinger, T.v.S., S.C., S.L.S. planned and carried out experiments or analyses. N.J.G., K.M.R., S.P., M.T., D.W., J.R.M., D.L.A., A.S., T.B., J.H.G., C.O., P.P.G., M. Muffato. M.P., K.B., F.J.M., B.P., L.K., P.M., T.R.B., M.W., Y.C., H.M., R.K.S., M.D.J., R.D.N., J.I.A., N.V., T.A.H., J.R., V.P., C.R.P., V.M.W., L.Z., R.D.K., J.M.R., V.V.Z., Z.W., D.S., M. Mariotti, R.G., S.M.R., V.G.T., N.D., H.T., H.A., J.M.P., D.G.M., V.L.G., C.G.B., D.P.D., S. Pabinger, S.L.S. contributed to the interpretation and presentation of results in the main manuscript and supplementary documents. N.J.G. wrote the first draft of the manuscript with input from all other authors.

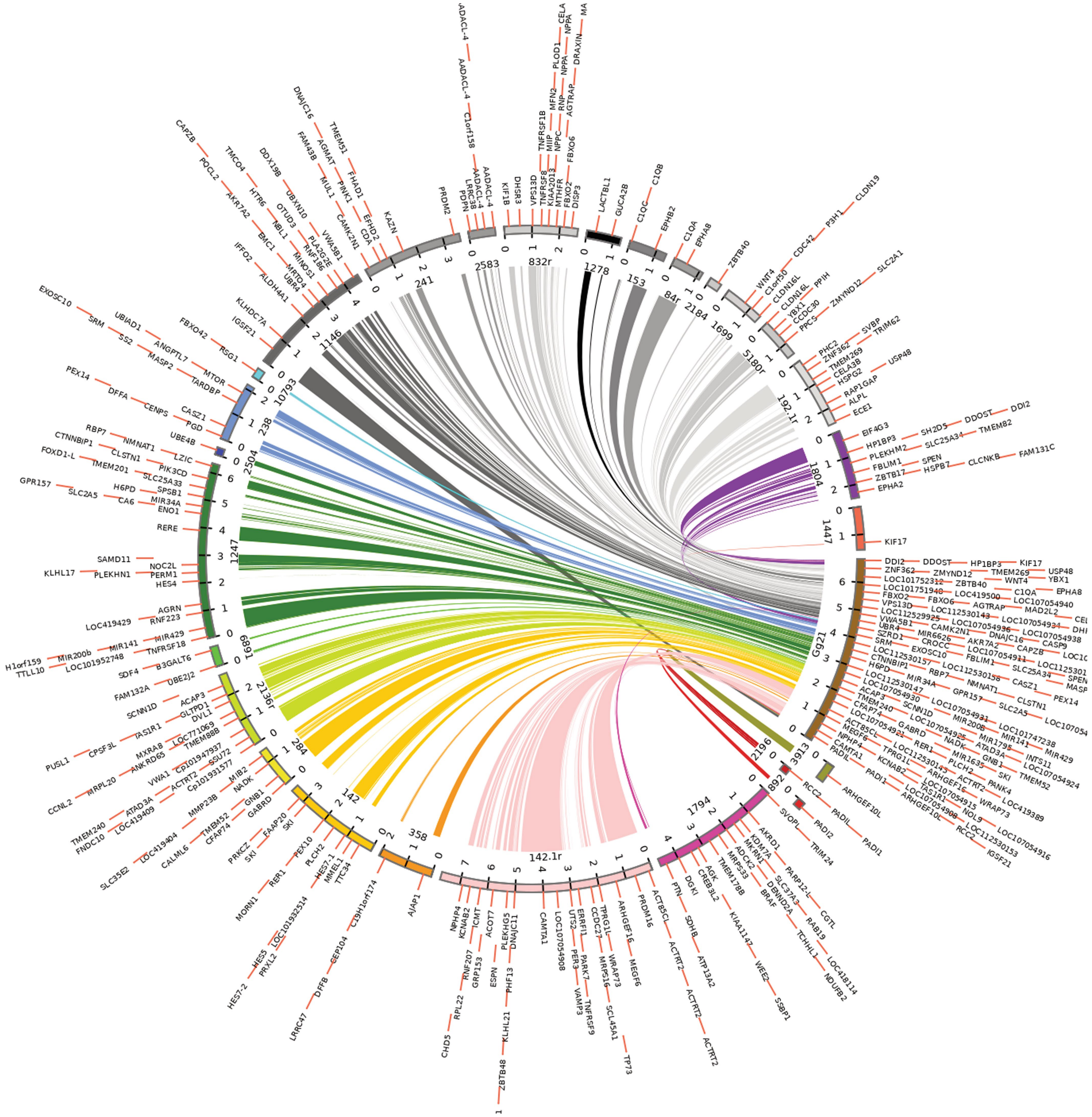
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2561-9.
Correspondence and requests for materials should be addressed to N.J.G.
Peer review information *Nature* thanks Rebecca Johnson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

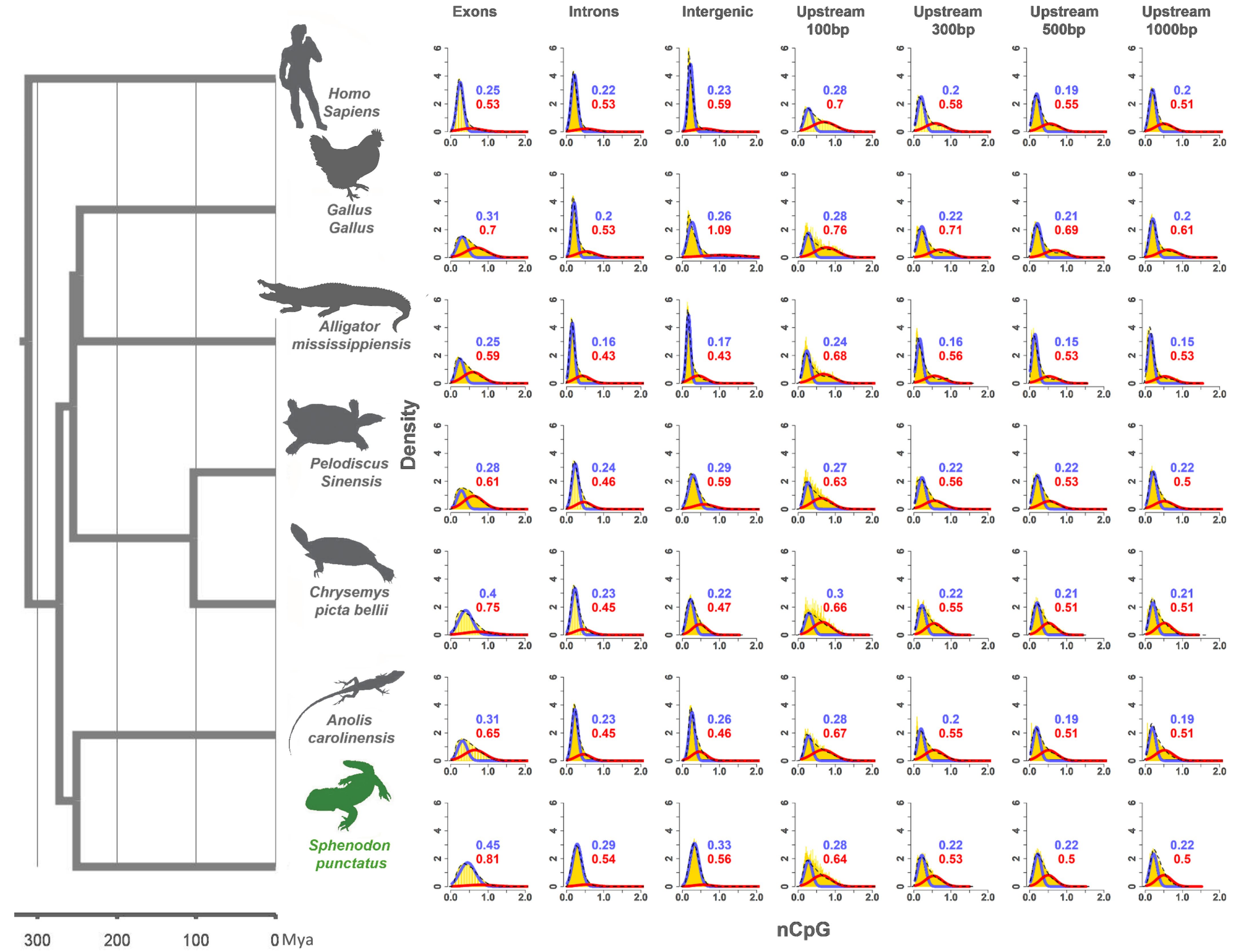


Extended Data Fig. 1 | See next page for caption.

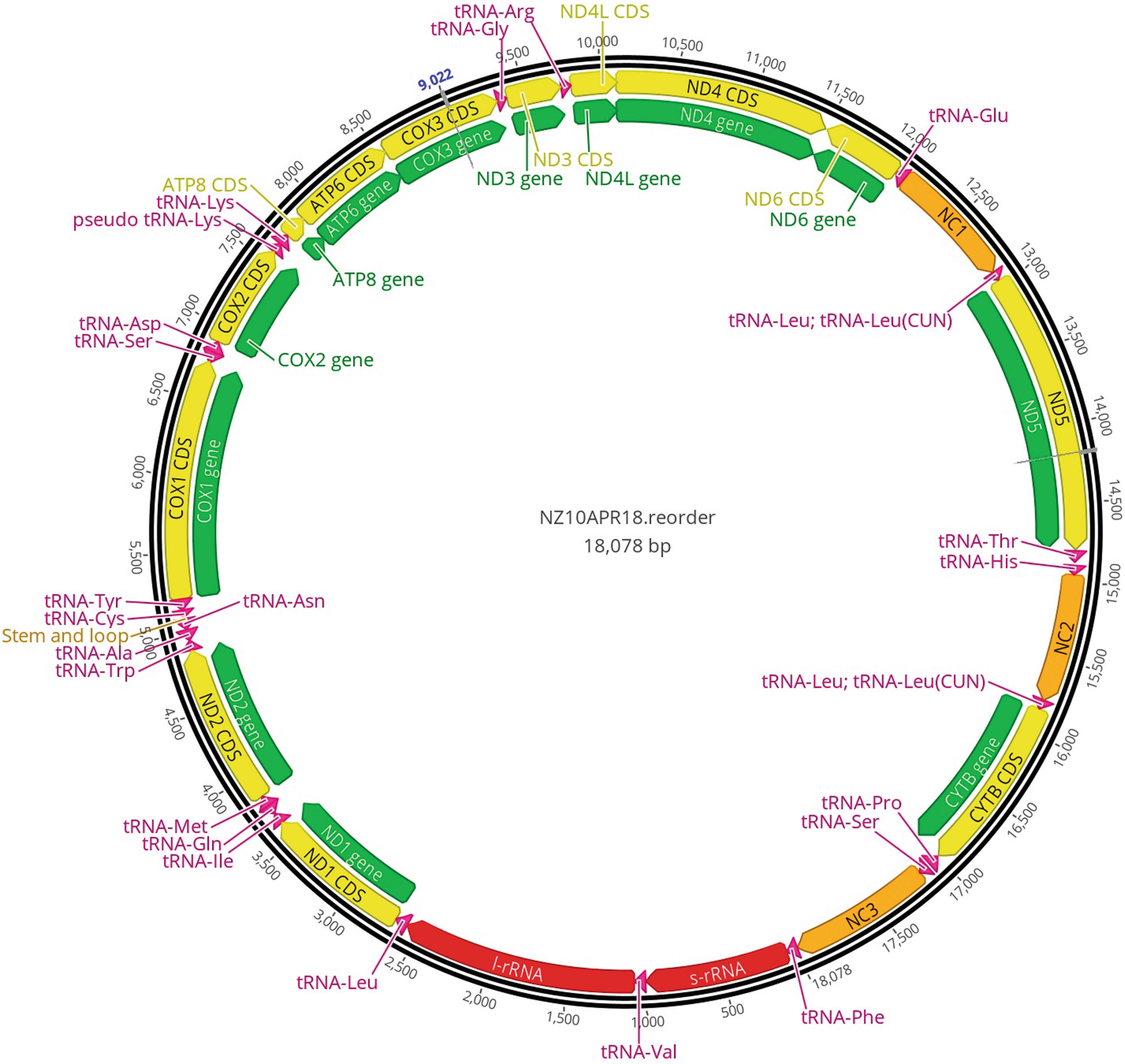
Extended Data Fig. 1 | Gene order conservation. a, Gene-order conservation score distribution using the tuatara as a reference. Species are ordered by the proportion of top-scoring orthologues ≥ 50. **b**, Gene-order conservation versus divergence time. For the three taxonomic groupings (Lepidosauria, Sauria and Amniota), we analysed the percentage of genes that are found in a conserved position across all pairs of genomes. Pairwise comparisons involving tuatara are shown in plain red circles (respectively, $n = 8$, $n = 10$, and $n = 4$), and comparisons that do not involve tuatara are black (box plot and + signs; respectively, $n = 0$, $n = 80$ and $n = 72$). The conservation of gene order between tuatara, birds and turtles is significantly higher (one-sided, two-sample Kolmogorov–Smirnov test, P value = 2.8×10^{-3}, $D = 0.575$) than that observed between squamates, birds and turtles. As the tuatara is the only remaining rhynchocephalian, there is no control distribution for the Lepidosauria ancestor. Box plot coordinates (minimum, first quartile, median, mean, third quartile, maximum) are 42.46%, 57.93%, 66.00%, 64.50%, 70.66% and 84.76% for the Sauria box plot, and 21.52%, 40.81%, 57.43%, 55.17%, 69.46% and 82.34% for the Amniota box plot.



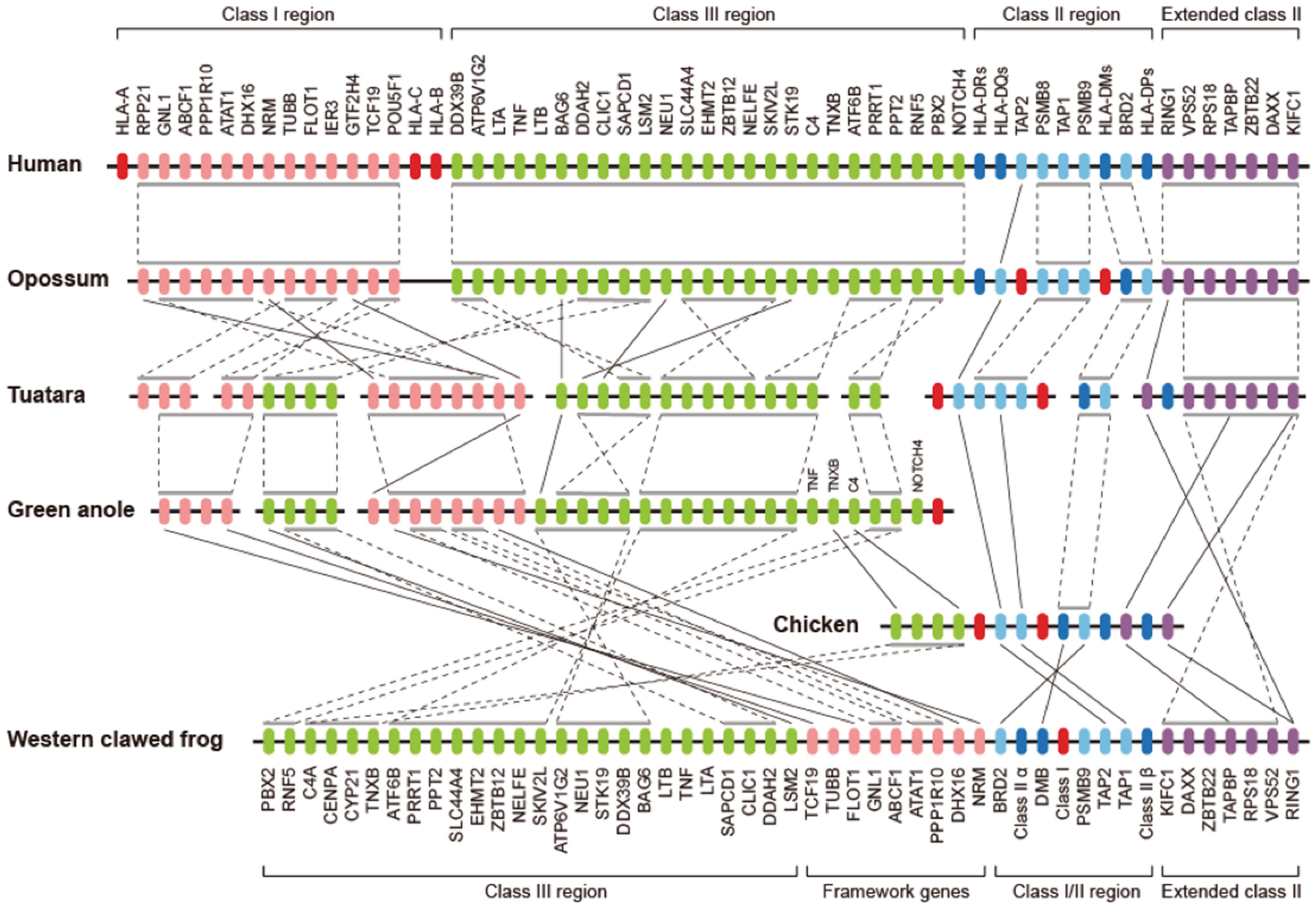
Extended Data Fig. 2 | Co-linearity between *Gallus gallus* chromosome 21 and tuatara contigs. Circos plot highlighting the gene-order conservation observed between chicken chromosome 21 (assembly GRCg6a) and multiple tuatara contigs. The gene names shown derive from the chicken assembly.



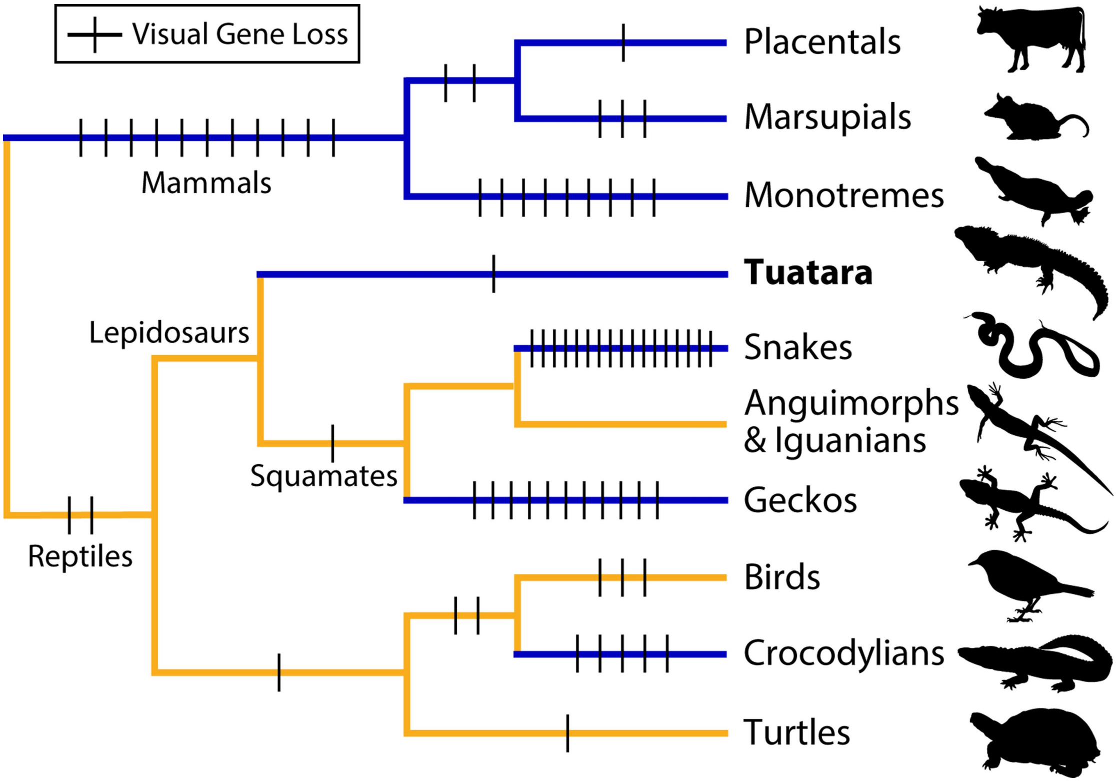
Extended Data Fig. 3 | Normalized CpG distributions for tuatara and other vertebrates. Tuatara show a statistically significant bimodal normalized CpG distribution in all genomic regions examined (gene promoters, exons, introns and intergenic sequences). Explicit testing using mixed models uncovered global normalized CpG bimodality across all taxa, indicating that bimodality is ancestral and highly conserved across vertebrates irrespective of sex-determining mechanism. Further details are provided in Supplementary Information 8. Silhouettes are hand-drawn from the original photographs of A. Eckley, except for *Anolis* (D. Hobern, CC-BY-2.0), crocodile and hen (designed by Creazilla, https://creazilla.com/).



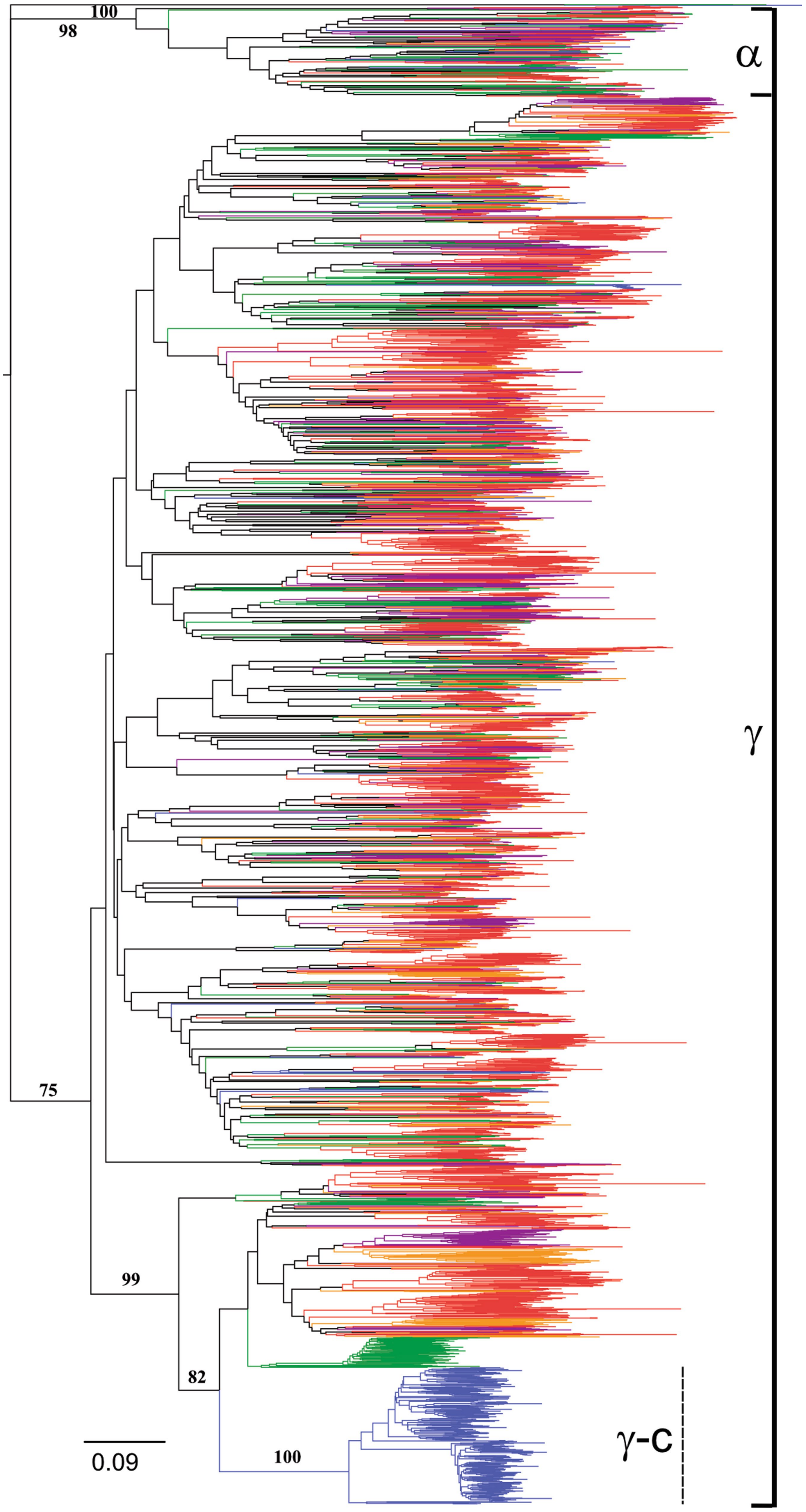
Extended Data Fig. 4 | The mitochondrial genome of the tuatara. In the Lady Alice Island reference tuatara, this molecule is 18,078 bp, containing 13 protein-coding, 2 rRNA and 22 tRNA genes, standard among animals and contradicting previous reports that 3 genes (*ND5*, t*RNA*^{Thr} and *tRNA*^{His}) were absent. Three non-coding areas with control region (heavy-strand replication origin) features (NC1, NC2 and NC3), and two copies of *tRNA*$^{Leu(CUN)}$ adjacent to NC1 and NC2 possess identical or near-identical sequence, possibly as a result of concerted evolution. A stem-and-loop structure is observed in the region encoding tRNA^{Asn} and tRNA^{Cys}, which may supplement for the origin of light-strand replication (O$_{L}$) normally found in this location. The *tRNA*^{Lys} gene is duplicated with the first copy, possibly a pseudo-gene. The *tRNA*^{Cys} gene encodes a tRNA with a D-arm replacement loop. The gene and structure order is: *tRNA*^{Phe}, *12S* rRNA, *tRNA*^{Val}, *16S* rRNA, *tRNA*$^{Leu(UUR)}$, ND1, *tRNA*^{Ile}, *tRNA*^{Gln}, *tRNA*^{Met}, ND2, *tRNA*^{Trp}, *tRNA*^{Ala}, *tRNA*^{Asn}, O$_{L}$-like structure, *tRNA*^{Cys}, *tRNA*^{Tyr}, *COI*, *tRNA*$^{Ser(UCN)}$, *tRNA*^{Asp}, *COII*, pseudo-*tRNA*^{Lys}, *tRNA*^{Lys}, *ATP8*, *ATP6*, *COIII*, *tRNA*^{Gly}, ND3, *tRNA*^{Arg}, *ND4L*, *ND4*, *ND6*, *tRNA*^{Glu}, NC1, *tRNA*$^{Leu(CUN)}$ copy one, *ND5*, *tRNA*^{Thr}, *tRNA*^{His}, NC2, *tRNA*$^{Leu(CUN)}$ copy two, *CYTB*, *tRNA*^{Pro}, *tRNA*$^{Ser(AGY)}$ and NC3.



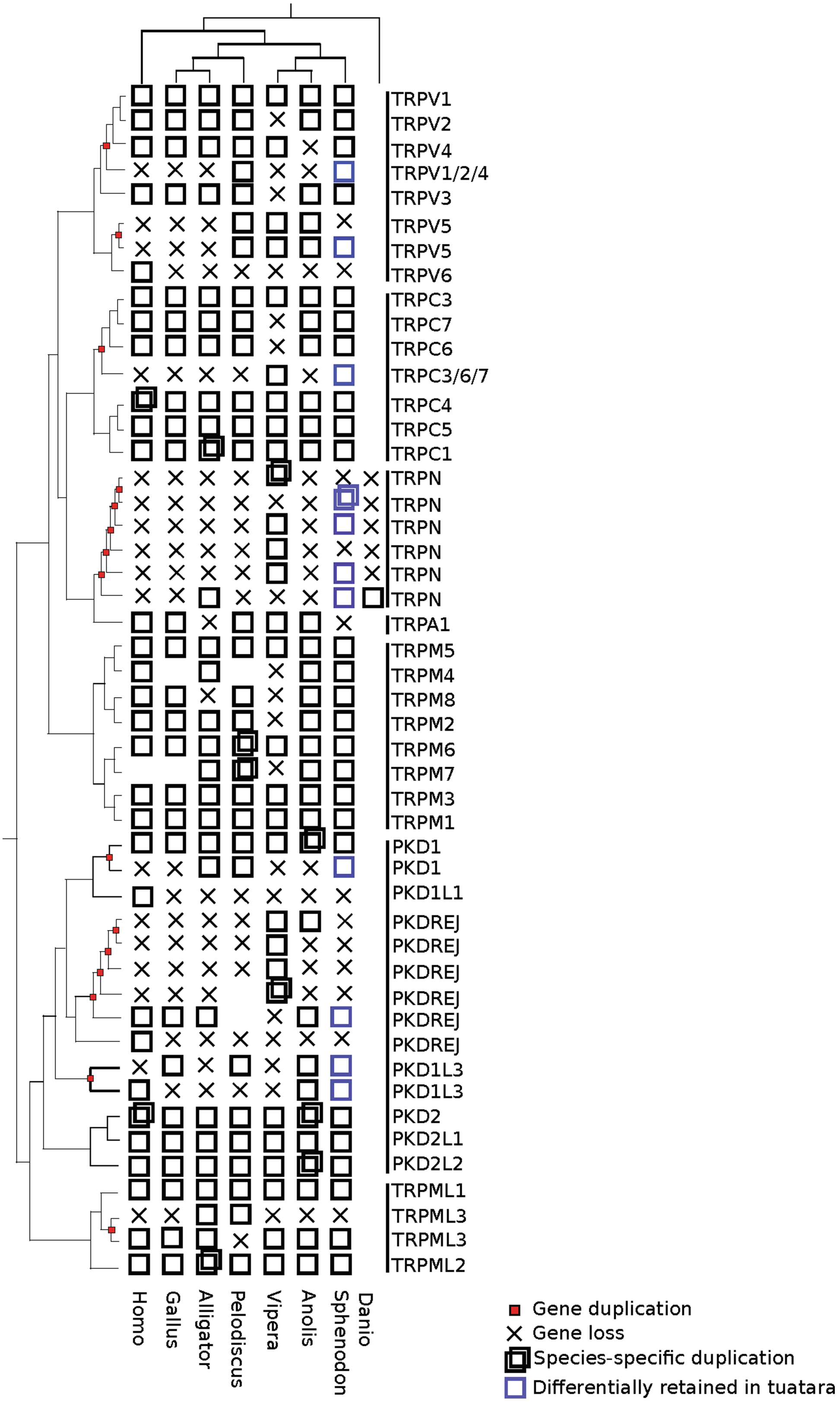
Extended Data Fig. 5 | Comparative analysis of the MHC core region. Only genes that were annotated in the tuatara genome were included in the analysis. Orthologues between species are connected by a solid line; the grey bars above and below genes indicate syntenic blocks and are linked by dashed lines between species. *Anolis* MHC class I/II and extended class II regions are not shown, owing to the high degree of genome assembly fragmentation in these regions. Red, class I genes; pink, class I region framework genes; green, class III genes; dark blue, class II genes; light blue, class II region framework genes; purple, extended class II genes.



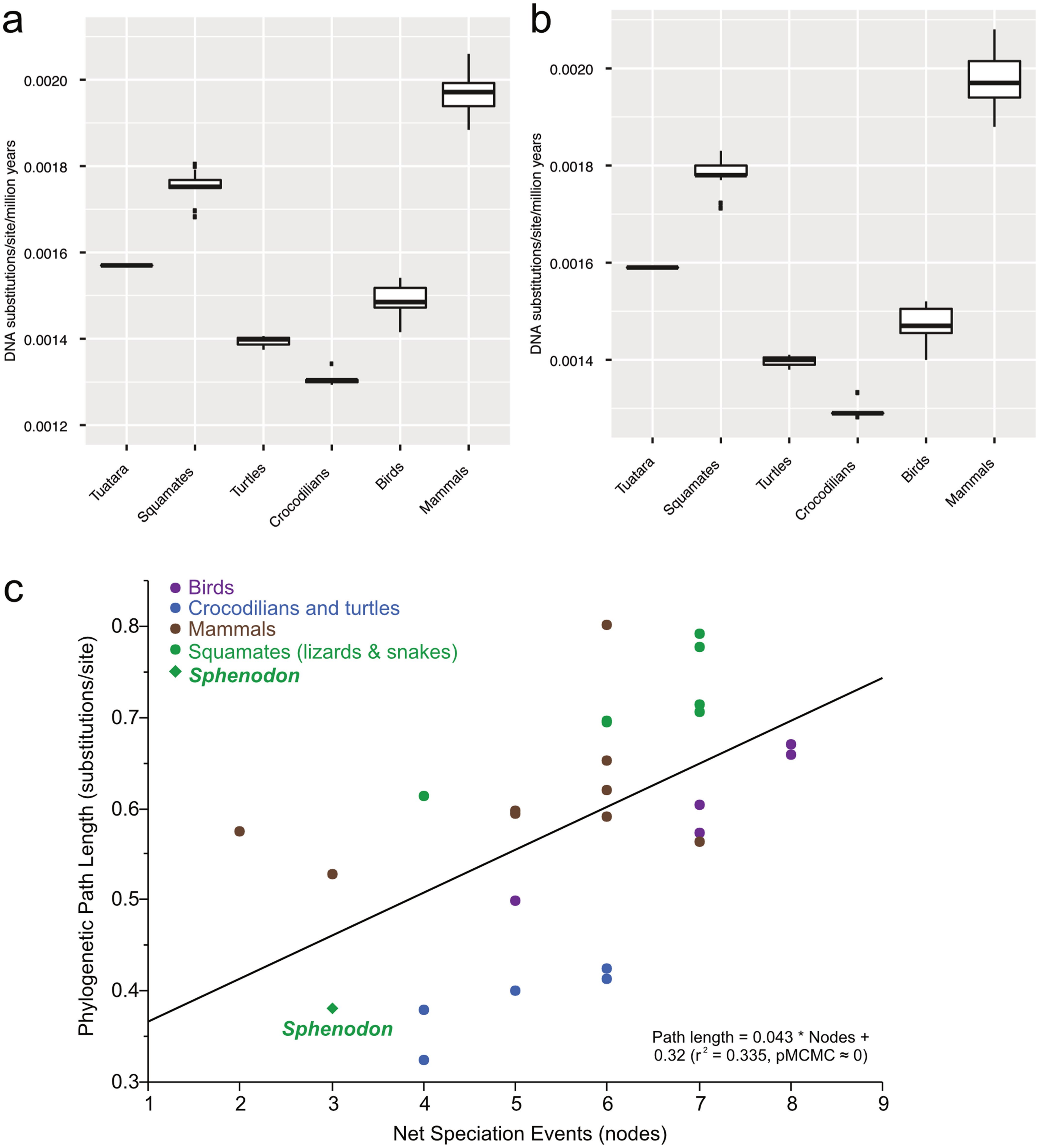
Extended Data Fig. 6 | Phylogenetic tree of amniotes depicting inferred visual gene losses. Lineages are coloured based on a rough approximation of their ancestral activity pattern (blue, nocturnal; yellow, diurnal). The tuatara lineage has experienced some of the lowest rates of gene loss despite a nocturnal ancestry, which in other lineages is associated with increased gene loss. Supplementary Table 10.1 provides a summary of genes lost per branch. The complete table of visual genes analysed and their presence or absence from the various groups is available at http://www.doi.org/10.5281/zenodo.2597599. Silhouettes were designed by Creazilla, www.creazilla.com.



Extended Data Fig. 7 | The evolutionary history of odorant receptors in terrestrial sauropsids. The relationship among odorant receptors was inferred using the neighbour-joining method. The unrooted tree contains 3,213 amino acid sequences. Branches are coloured according to the following categories: green, tuatara; blue, birds (*G. gallus* and *Taeniopygia guttata*); red, snakes (*Notechis scutatus*, *Ophiophagus hannah, Protobothrops mucrosquamatus*, *Pseudonaja textilis*, *P. bivittatus* and *Thamnophis sirtalis*); orange, lizards (*A. carolinensis* and *Pogona vitticeps*); and purple, gecko (*Gekko japonicas*). Bootstrap support values above 75% (1,000 replicates) are indicated for major branch splits relating to the different odorant receptor groups and branches leading to the species-specific odorant receptor expansions in birds (group γ−c) and tuatara (*).



Extended Data Fig. 8 | The repertoire of the TRP genes identified in tuatara. We compared the TRP genes identified in tuatara (*S. punctatus)* to those found in other six vertebrate species: lizard (*A. carolinensis*), viper (*Vipera berus*), turtle (*Pelodiscus sinensis*), alligator (*Alligator mississippiensis*), chicken (*G. gallus*), and human (*Homo sapiens*). Small red squares on nodes indicate gene duplications, duplicated boxes indicate species-specific duplications and blue boxes indicate differential gene retentions in tuatara. Crosses indicate genes that were lost after duplication and empty spaces genes that have most probably been lost, but are not yet confirmed.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Estimated DNA substitution rates of amniote clades on the basis of fourfold-degenerate sites and a test for evidence of punctuated evolution. a, Box plots showing distribution of estimated substitution rates by clade using semiparametric penalized likelihood in r8s with fossil constraints from ref. [36] and ref. [37] for tuatara ($n = 1$, 0.00159), squamates ($n = 13$, minimum = 0.00171, maximum = 0.00183, median = 0.00178, 25th percentile = 0.00178, 75th percentile = 0.00183), turtles ($n = 3$, minimum = 0.00138, maximum = 0.00141, median = 0.00140, 25th percentile = 0.00139, 75th percentile = 0.00140), crocodilians ($n = 5$, minimum = 0.00128, maximum = 0.0133, median = 0.00129, 25th percentile = 0.00129, 75th percentile = 0.00129), birds ($n = 11$, minimum = 0.0014, maximum = 0.00152, median = 0.00147, 25th percentile = 0.00146, 75th percentile = 0.0015) and mammals ($n = 15$, minimum = 0.00188, maximum = 0.00208, median = 0.00197, 25th percentile = 0.00194, 75th percentile = 0.00201). **b**, Box plots showing distribution of estimated substitution rates with median time to most recent common ancestor estimates from www.timetree.org for tuatara ($n = 1$, 0.00157), squamates ($n = 13$, minimum = 0.00168, maximum = 0.00180, median = 0.00175, 25th percentile = 0.00178, 75th percentile = 0.00177), turtles ($n = 3$, minimum = 0.00138, maximum = 0.00141, median = 0.00140, 25th percentile = 0.00139, 75th percentile = 0.00140), crocodilians ($n = 5$, minimum = 0.00129, maximum = 0.0134, median = 0.00130, 25th percentile = 0.00130, 75th percentile = 0.00130), birds ($n = 11$, minimum = 0.00142, maximum = 0.00154, median = 0.00149, 25th percentile = 0.00147, 75th percentile = 0.00152) and mammals ($n = 15$, minimum = 0.00188, maximum = 0.00206, median = 0.00197, 25th percentile = 0.00194, 75th percentile = 0.00199). **c**, A test for evidence of punctuated evolution. The process of punctuated genome evolution predicts that the amount of evolution in the genome of a given species should correlate with the net number of speciation events. We used Bayesian phylogenetic generalized least squares to regress the total phylogenetic path length (of fourfold-degenerate sites) on the net number of speciation events (nodes in a phylogenetic tree). We find strong evidence for punctuated evolution, which accounts for 33.5% (r^2; 95% credible interval = 0.34 to 0.38) of deviation from the molecular clock at fourfold-degenerate sites.

Extended Data Table 1 | Assembly statistics and quality metrics for the tuatara genome

	ALLPATHS-LG assembly	Final assembly
Number of scaffolds	48545	16537
Total size of scaffolds	4338416404	4272217537
Longest scaffold	6644062	29987930
N50 scaffold length	322768	3052611
Number of contigs	580919	270321
Longest contig	141552	347348
N50 contig length	10517	27421
CEGMA complete count	89	98
CEGMA complete %	35.89%	39.52%
CEGMA partial count	209	210
CEGMA partial %	84.27%	84.68%
Complete vertebrata BUSCOs	2688	2911
Complete %	80.1	86.8
Complete and single-copy BUSCOs	2667	2882
Complete and duplicated BUSCOs	21	29
Fragmented BUSCOs	353	218
Missing BUSCOs	313	225
Total BUSCO groups searched	3354	3354

nature research

Corresponding author(s): Neil J. Gemmell et al.

Last updated by author(s): 4/5/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Standard genome assembly and bioinformatic tools were employed and are detailed in full in the methods and supplementary materials for the paper. Genome assembly was undertaken using AllPaths-LG (version 49856, http://software.broadinstitute.org/allpaths-lg/blog/?page_id=12) and Dovetail Genomics HiRise scaffolding software. Transcriptomes were assembled using Trinity v2.2.0. Bisuplhite sequencing data were trimmed using TrimGalore v0.4.0 and reads mapped using Bismark v0.14.311 to identify metylated sites. Repeat annotation was undertaken using CARP, RepeatModeler and LTRharvest. Gene annotation used RepeatMasker (4.0.3) and MAKER2. Genotype-by-sequencing was undertaken using FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by a QC analysis pipeline: "Deconvolute and quality control" https://github.com/AgResearch/DECONVQC] and subsequent demultiplexing using GBSX, read mapping using BWA MEM, and SNV calling using STACKS and GATK.
Data analysis	Standard bioinformatic tools were employed for our analyses. These and are detailed in the methods and supplementary materials for the paper. Where custom code was utilized this is also specified and either available from GitHub or directly from the authors of the relevant section of our manuscript. All attributions to each component of our work are clearly signalled. For completeness the full list is provided here also: Repeat and gene annotation RepeatMasker (v4.0.5), http://www.repeatmasker.org/ MAKER2 (v2.31.8), http://www.yandell-lab.org/software/maker.html BLAST (v2.2.30+), https://blast.ncbi.nlm.nih.gov/Blast.cgi SNAP (v2.4.7), http://snap.cs.berkeley.edu Augustus (v3.3), http://augustus.gobics.de BUSCO (v3.0), https://busco.ezlab.org Ortholog calling

Ensembl GeneTree pipeline: https://github.com/Ensembl/ensembl and https://github.com/Ensembl/ensembl-compara branch “release/87”
Ensembl Hive workflow management system https://github.com/Ensembl/ensembl-hive branch “version/2.3”
Plotting script: https://github.com/Ensembl/ensembl-compara/blob/release/89/scripts/homology/plotGocData.r
hcluster 0.5.0 https://sourceforge.net/projects/treesoft
TCoffee 9.03.r1318 http://www.tcoffee.org/
Mafft 7.221 https://mafft.cbrc.jp/alignment/software/mafft-7.221-with-extensions-src.tgz
Treebest https://github.com/Ensembl/treebest
QuickTree 1.1 https://github.com/khowe/quicktree
NCBI Blast 2.2.30+ ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.30/
HMMER 2.3.2 http://eddylab.org/software/hmmer/2.3.2/hmmer-2.3.2.tar.gz
PantherScore 1.0.3 https://github.com/Ensembl/pantherScore
PAML 4.3 http://abacus.gene.ucl.ac.uk/software/SoftOld/paml4.3.tar.gz
Ktreedist 1.0.0 http://molevol.cmima.csic.es/castresana/Ktreedist/Ktreedist_v1.tar.gz
CAFE 2.2 http://downloads.sourceforge.net/project/cafehahnlab/Previous_versions/cafehahnlab-code_v2.2.tgz

Ensembl annotation
RepeatMasker, http://repeatmasker.systemsbiology.net/
RepeatModeler, http://www.repeatmasker.org/RepeatModeler/
GenBlast, http://genome.sfu.ca/genblast/
BWA, http://bio-bwa.sourceforge.net/
MUSCLE, http://www.drive5.com/muscle
BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi
ensembl-analysis, https://github.com/Ensembl/ensembl-analysis.git
ensembl code, https://www.ensembl.org/info/docs/api/api_installation.html
exonerate, https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
rnafold, https://github.com/choener/RNAfold
Inferno, http://eddylab.org/infernal/

Investigation of gene co-linearity
BLAST+ 2.8.1 using Geneious version 10.2.6 (https://www.geneious.com).
Circos software http://circos.ca/software/

Ab initio repeat annotation
CARP, https://github.com/carp-te/carp-documentation RepeatModeler, http://www.repeatmasker.org/RepeatModeler/
CENSOR, which requires wu-blast and bioperl, https://girinst.org/downloads/software/censor/ RepeatMasker, http://www.repeatmasker.org/
BLASTN, https://blast.ncbi.nlm.nih.gov/Blast.cgi RepBase, https://www.girinst.org/server/RepBase/
MUSCLE, https://www.drive5.com/muscle/
MrBayes, https://nbisweden.github.io/MrBayes/download.html
FastTree, http://www.microbesonline.org/fasttree/
USEARCH, https://www.drive5.com/usearch/
HMMER, http://hmmer.org/ PILER, https://www.drive5.com/piler/

Repeat annotation of SINEs and DNA transposons
RepeatModeler version 1.0.8 http://www.repeatmasker.org/RepeatModeler/
renameRMDLconsensi.pl script https://github.com/genomicrocosm/physaliaTEcourse/blob/master/Practical2_Computational_annotation/renameRMDLconsensi.pl
repeatModelerPipeline4.pl script https://github.com/genomicrocosm/physaliaTEcourse/blob/master/Practical3_Manual_curation/repeatModelerPipeline4.pl
BLASTn version 2.2.28+ https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download
MAFFT version 6 standalone and version 7 webserver https://mafft.cbrc.jp/alignment/software/
BioEdit version 7.2.6.1 https://bioedit.software.informer.com/download/
CENSOR webserver http://www.girinst.org/censor/index.php
RAxML version 8.0.0 in CIPRES Science Gateway https://www.phylo.org/portal2/login!input.action
MEGA version 5.2 https://www.megasoftware.net/
RepeatMasker version 4.0.7 http://www.repeatmasker.org/RMDownload.html
calcDivergenceFromAlign.pl script in RepeatMasker package http://www.repeatmasker.org/RMDownload.html
R scripts for making landscape plots https://github.com/ValentinaBoP/TuataraTELandscapes/blob/master/Tuatara_DNA_SINE_landscape_figures.Rmd

LTR analyses
Genometools, http://genometools.org/pub/genometools-1.5.8.tar.gz, used for indexing genome and running LTRharvest.
Blastn, https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.5.0/ncbi-blast-2.5.0+-x64-linux.tar.gz, used for blasting retroviral proteins.
MAFFT 7, https://mafft.cbrc.jp/alignment/software/, used for multiple alignment of peptide sequences.
CD-HIT-V4.6.5, https://github.com/weizhongli/cdhit, used to reduce number of similar retroelement copies in file.
RAxML 8.2, https://github.com/stamatak/standard-RAxML, used to infer phylogenetic trees under maximum likelihood.
NINJA 1.2.2, http://nimbletwist.com/software/ninja/old_distros/ninja_1.2.2.tgz, used for large-scale neighbor-joining phylogeny inference.
Inkscape 0.92, https://inkscape.org/release/inkscape-0.92/, used for preparing figures of repetitive elements.

RNA annotation
Rfam (version 13.0) covariance models, ftp://ftp.ebi.ac.uk/pub/databases/Rfam/13.0/
tRNAscan-SE (version 1.3.1), http://lowelab.ucsc.edu/software/
Infernal (version 1.1), http://eddylab.org/infernal/

Mitochondrial genome sequence and assembly
Illumina data: Bowtie 2, MaSuRCA, Minimus, Jellyfish, Geneious v10.2.4 (https://www.geneious.com).
Oxford Nanopore data: Nanopolish, Guppy, Blastn, Megablast, Discontiguous Megablast.

MHC
BLAST+ 2.3.0 https://blast.ncbi.nlm.nih.gov/Blast.cgi Geneious 9.1, (https://www.geneious.com).

Opsin gene analysis
BLAST+ 2.2.30, PAML 4.8, MEGA 5.2, PhyML 3.0, BLASTPHYME https://github.com/ryankschott/BlastPhyMe

Odorant receptors
tBLASTn in Geneious 10.0.3 (https://www.geneious.com), MAFFT (v7.338), MEGA 7.0.21, FigTree v1.4.4

Transient receptor potential (TRP) ion channel gene analysis
tBLASTn, https://blast.ncbi.nlm.nih.gov/Blast.cgi MAFFT v7.450 / https://mafft.cbrc.jp/alignment/software/
FastTree2 as implemented in the CIPRES portal https://www.phylo.org/

Selenoprotein analysis
Selenoprofiles v. 3.6, https://github.com/marco-mariotti/selenoprofiles
Secmarker v. 0.4, https://secmarker.crg.cat,

Phylogeny and evolutionary rates
LASTZ-chaining-netting/source code from the avian phylogenomics project (Zhang et al. 2014; GigaScience) https://github.com/gigascience/paper-zhang2014/tree/master/Whole_genome_alignment/
msa_view in PHAST v1.3 /http://compgen.cshl.edu/phast/oldversions.php/
RAxML v8.2.3/https://github.com/stamatak/standard-RAxML
phyloFit in PHAST v1.3 /http://compgen.cshl.edu/phast/oldversions.php/
r8s v1.8 / https://sourceforge.net/projects/r8s/
PRANK v1.7 / http://wasabiapp.org/download/prank/
FASconCAT-G v.1.02 / https://github.com/PatrickKueck/FASconCAT-G
RAxML v8.2.3/https://github.com/stamatak/standard-RAxML
ASTRAL v4 / https://github.com/smirarab/ASTRAL
MAFFT v7.450 / https://mafft.cbrc.jp/alignment/software/
AMAS / https://github.com/marekborowiec/AMAS
IQ-TREE v1.6.12 / http://www.iqtree.org/

Punctuated evolution
Time-calibrated phylogeny: BEAST v2.4.8 on the CIPRES Science Gateway: https://www.phylo.org/
Punctuated evolution: BayesTraits V3.0.2 Nov 2019: http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.2/BayesTraitsV3.0.2.html
Node-density artefact: Test for Punctuational Evolution and the Node-Density Artifact. v1: http://www.evolution.reading.ac.uk/pe/index.html

Patterns of selection
Translatorx, http://translatorx.co.uk/, Translation of nucleotide sequences
MAFFT 7.310, https://mafft.cbrc.jp/alignment/software/ , amino acid alignment
PAL2NALv14, https://github.com/drostlab/orthologr/tree/master/inst/pal2nal/pal2nal.v14 , conversion of protein alignments into codon-based DNA alignments
TrimAl 1.2, http://trimal.cgenomics.org/, Alignment correction
Garli 2.0.1, https://www.nescent.org/wg_garli/Main_Page, maximum likelihood phylogenetic reconstruction
PAML 4.5, http://abacus.gene.ucl.ac.uk/software/paml.html , infer branch specific evolutionary patterns
QVALUE, https://www.bioconductor.org/packages/release/bioc/html/qvalue.html , false discovery rate correction
Kobas 2.0, http://kobas.cbi.pku.edu.cn/kobas3/?t=1, Gene set enrichment analysis
Galaxy's Stitch Gene Blocks Tool, http://www.bioinformatics.nl/galaxy
AlignmentProcessor0.12, https://github.com/WilsonSayresLab/AlignmentProcessor
PAML, CodeML, http://abacus.gene.ucl.ac.uk/software/paml.html
R v3.3.1, https://cran.r-project.org/

Reconstruction of the demographic history of the tuatara
BWA mem, http://bio-bwa.sourceforge.net/
Samtools mpileup, http://www.htslib.org/
PSMC, https://github.com/lh3/psmc

Population genomics
FastQC v0.10.1, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ QC analysis pipeline, https://github.com/AgResearch/DECONVQC
GBSX, https://github.com/GenomicsCoreLeuven/GBSX
BWA mem, http://bio-bwa.sourceforge.net/
STACKS 1.4.4, http://catchenlab.life.illinois.edu/stacks/
GATK haplotypecaller, https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data are freely available. The Tuatara Genome Consortium Project Whole Genome Shotgun and genome assembly are registered under the umbrella BioProject PRJNA418887 and BioSample SAMN08038466. Transcriptome read data are submitted under SRR7084910 (whole blood) together with prior data SRR485948. The transcriptome assembly is submitted to GenBank with ID GGNQ00000000.1. Illumina short-read and nanopore long read sequence are in SRAs associated with PRJNA445603. The assembly (GCA_003113815.1) described in this paper is version QEPC00000000.1 and consists of sequences QEPC01000001-QEPC01016536. Maker gene predictions are available from Zenodo, DOI: 10.5281/zenodo.1489353. The repeat library database developed for tuatara is available from Zenodo, DOI: 10.5281/zenodo.2585367. Other data for analyses in specfic sections of our paper have been uploaded to Zenodo and DOIs are clearly indicated in the paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is the study of the genome, transcriptome and methylome of one exemplar male tuatara from Lady Alice Island in the far North of New Zealand. We have subsequently compared this genome to other published genomes, while also undertaking population geneomic investigations of tuatara using 30 individuals that span the main axes of genetic diversity for tuatara .
Research sample	1 exemplar tuatara (Sphenodon punctatus) from Lady Alice Is. New Zealand. 30 other tuatara samples (approximately half male, half female) spanning three main populations (Lady Alice, Stephens, and Brothers Is.) that encompass the main axes of genetic diversity for this species.
Sampling strategy	Sampling was undertaken using venipuncture. Because this species is special to Maori and highly protected, our sampling strategy was ad hoc relying heavily on samples collected previously for other studies. While no tests of statistical power were undertaken past investigations of population structure in this species suggested that 10 samples per population was likely adequate to capture much of the variation present. The use of equal numbers of males and females provided a reasonable opportunity to explore obvious sex differences, should these be present.
Data collection	Blood samples for the population-level analysis were collected during field work for research on other projects, for example, investigating population size and genetic diversity of island populations of tuatara. Islands were searched for tuatara at night when they were active above ground. Tuatara were captured by hand, and ~1ml blood samples were obtained from the caudal artery. Samples were stored in liquid nitrogen then once back at the lab, transferred to a -80 ˚C freezer. The exemplar tuatara sample was collected during a survey trip to Lady Alice Island, but otherwise procedures were similar to earlier population samples.
Timing and spatial scale	Samples were collected at each of three sites (Lady Alice, Stephens, and Brothers Is.) from 1984 to 2011.
Data exclusions	Several samples (2 out of 30) failed in downstream genotype by sequencing and bisulphite methylation profiling due to DNA quality/ qunatity issues.
Reproducibility	Data reproducibility was verified using repeat sequencing and independent analyses using alternative pipelines. e.g. genome assembly used at least three independent pipelines, each of which had high concordance.
Randomization	There was no need to randomise our study given the focus on genomic and population genomics
Blinding	Blinding was irrelevant given the focus on genomic and population genomics

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Data not available, but field sites are all temperate offshore Islands in New Zealand
Location	Lady Alice, Stephens, and Brothers Is. New Zealand
Access and import/export	Samples were collected with the permission and support of local iwi and the NZ DOC, under Victoria University of Wellington

Access and import/export	Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347 and held and used under permits 45462-DOA (1/9/15) and 32037-RES (11/11/11) issued by the New Zealand Department of Conservation. Samples from the exemplar Lady Alice animals were shipped from NZ to ZSD in USA using CITES Permit to Export – Permit # 13NZ000096 (25/7/13), NZ Dept. of Conservation Authority to Export – Permit # 36830-RES (11/7/2013), and CITES Import Permit # 13US727416/9 and US Federal Fish and Wildlife Permit # LE736007-0 (15/7/13). The ethics application and other permitting processes ensures minimal numbers of individuals are used in research, and that their use is justified under New Zealand and international law. All approvals issued by the Department of Conservation ensure the research complies with relevant acts of Parliament for access to collection sites and handing and research on native species of New Zealand. All Department of Conservation permits for capture, sampling, require consultation with local indigenous people affiliated with the islands.
Disturbance	Animals were handled minimally and returned to the site of capture for release.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	This study did not involve the us of laboratory animals
Wild animals	Adult tuatara, (Sphenodon punctatus) from Lady Alice, Stephens, and Brothers Is. New Zealand were captured and blood samples taken using established venipuncture approaches. Animals were captured while active outside their burrows at night. Blood samples were taken upon capture and animals were released at their site of capture. If animals were held while others were being sampled (<1h), they were placed into a cloth bag. No animals died as a result of this study.
Field-collected samples	Bloods from one exemplar male and a further 30 animals, previously collected for other purposes, were utlised for our work. Sex ratios among samples were approximately 50:50 and equal numbers of samples were obtained from all sites. One sample was collected specifically for this study - the exemplar - but during another research project, so it did not require separate arrangements solely for this sample. All others arose from frozen samples from previous studies.
Ethics oversight	Samples were collected under Victoria University of Wellington Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Article

Extracellular proteostasis prevents aggregation during pathogenic attack

https://doi.org/10.1038/s41586-020-2461-z

Received: 18 October 2018

Accepted: 16 April 2020

Ivan Gallotta[1], Aneet Sandhu[1,2], Maximilian Peters[3], Martin Haslbeck[4], Raimund Jung[1], Sinem Agilkaya[1], Jane L. Blersch[1,2], Christian Rödelsperger[5], Waltraud Röseler[5], Chaolie Huang[1], Ralf J. Sommer[5] & Della C. David[1,6] ✉

In metazoans, the secreted proteome participates in intercellular signalling and innate immunity, and builds the extracellular matrix scaffold around cells. Compared with the relatively constant intracellular environment, conditions for proteins in the extracellular space are harsher, and low concentrations of ATP prevent the activity of intracellular components of the protein quality-control machinery. Until now, only a few bona fide extracellular chaperones and proteases have been shown to limit the aggregation of extracellular proteins[1–5]. Here we performed a systematic analysis of the extracellular proteostasis network in *Caenorhabditis elegans* with an RNA interference screen that targets genes that encode the secreted proteome. We discovered 57 regulators of extracellular protein aggregation, including several proteins related to innate immunity. Because intracellular proteostasis is upregulated in response to pathogens[6–9], we investigated whether pathogens also stimulate extracellular proteostasis. Using a pore-forming toxin to mimic a pathogenic attack, we found that *C. elegans* responded by increasing the expression of components of extracellular proteostasis and by limiting aggregation of extracellular proteins. The activation of extracellular proteostasis was dependent on stress-activated MAP kinase signalling. Notably, the overexpression of components of extracellular proteostasis delayed ageing and rendered worms resistant to intoxication. We propose that enhanced extracellular proteostasis contributes to systemic host defence by maintaining a functional secreted proteome and avoiding proteotoxicity.

Extracellular pathological deposits are associated with a variety of diseases such as Alzheimer's disease, spongiform encephalopathies, cardiac amyloidosis and type II diabetes. A better understanding of the regulation of protein homeostasis (proteostasis) in the extracellular space could ultimately expand treatment options. The cellular proteostasis network comprises more than 2,000 factors[10], yet these factors are largely inactive outside the cell. In support of an active extracellular proteostasis network, a growing number of extracellular chaperones[1–3] have been identified to act as suppressors of aggregation by binding to misfolded proteins or oligomers and promoting their removal through receptor-mediated endocytosis. One of the main obstacles delaying the exploration of extracellular proteostasis in vivo is the lack of an amenable model for a comprehensive study. Here, we took advantage of the simplicity of *C. elegans*, in which the extracellular fluid in the body cavity (pseudocoelom) bathes all internal organs and provides a medium to exchange intercellular signals and distribute nutrients[11]. Six scavenger cells (coelomocytes) ensure turnover of extracellular components by non-specific endocytosis[12]. Except for this basic control by the coelomocytes, to our knowledge, nothing is known about an extracellular quality-control system for damaged proteins.

To discover extracellular regulators (ECRs), we constructed a *C. elegans* model to follow protein aggregation in the extracellular space. A previous proteomic analysis of the *C. elegans* aggregating proteome identified several secreted proteins highly prone to aggregate with age[13]. One of these extracellular proteins was lipid-binding protein 2 (LBP-2). Of note, mutations linked to Charcot–Marie–Tooth in the closest orthologue of LBP-2 in humans, fatty acid-binding protein myelin P2 (E value = 9.5×10^{-9}), increase its aggregation propensity[14]. LBP-2 labelled with the fluorescent protein tagRFP is secreted by body-wall muscles and is diffusely localized in the pseudocoelom in young animals and taken up by the coelomocytes (Fig. 1a–c, Extended Data Fig. 1a). Toxin-mediated ablation of coelomocytes causes LBP-2 to accumulate with a similar distribution to GFP secreted from the body-wall muscles[12] (Extended Data Fig. 1b, c). As animals aged and in young animals lacking coelomocytes, we observed the formation of LBP-2 puncta localized in the same space as secreted GFP outside of the body-wall muscles or neurons, consistent with its aggregation in the pseudocoelom (Fig. 1d–i, Extended Data Fig. 1d–f, Supplementary Video 1). Increased formation of LBP-2 puncta correlated with an increase in detergent-insoluble LBP-2, a hallmark of age-dependent protein aggregation and disease-associated protein aggregation (Fig. 1j). By contrast,

---

[1]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany. [2]Graduate Training Centre of Neuroscience, International Max Planck Research School, University of Tübingen, Tübingen, Germany. [3]Department of Medical Neurobiology, Hebrew University of Jerusalem, Jerusalem, Israel. [4]Department of Chemistry, Technical University of Munich, Garching, Germany. [5]Max Planck Institute for Developmental Biology, Department for Integrative Evolutionary Biology, Tübingen, Germany. [6]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany. ✉e-mail: della-crystal.david@uni-tuebingen.de

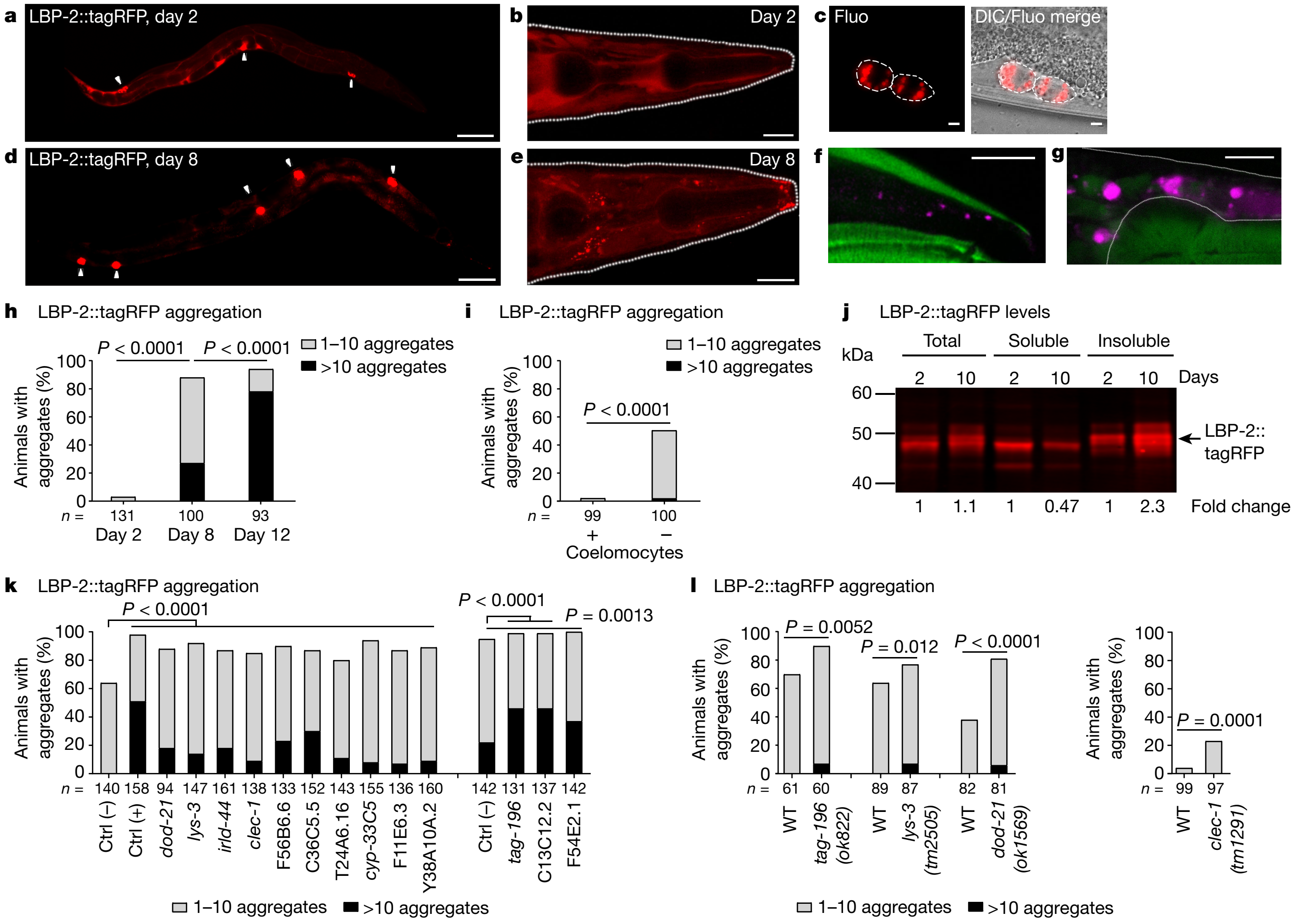


Fig. 1 | Secreted LBP-2 aggregates in the extracellular space with age in *C. elegans*. **a**, **b**, LBP-2::tagRFP in young animal (representative image of *n* = 16 worms) (**a**) and diffuse localization in head region (*n* = 24 worms) (**b**). **c**, LBP-2::tagRFP in coelomocytes (dashed circles, *n* = 5 worms). **d**, **e**, LBP-2::tagRFP in aged animal (*n* = 8 worms) (**d**) and puncta localized in head region (*n* = 27 worms) (**e**). Images in **a** and **d** are from identical exposures; arrowheads mark coelomocytes. Images in **b** and **e** are maximum projections, laser intensity 10% (**b**) and 8% (**e**). **f**, LBP-2::tagRFP puncta formation (magenta) between body-wall muscle (top) and pharyngeal muscle (bottom) (F-actin, green) (day 4, *n* = 7 worms). **g**, Without coelomocytes, LBP-2::tagRFP accumulates as puncta (magenta) surrounded by secreted GFP (green) adjacent to pharynx (GFP pharyngeal reporter, lower outline) (day 6, *n* = 5 worms). Images in **f** and **g** are from a single plane. Scale bars, 100 µm (**a**, **d**), 5 µm (**c**), 20 µm (**b**, **e**) or 10 µm (**f**, **g**). **h**, **i**, Quantification of LBP-2::tagRFP aggregation with age (*n* = 2 independent experiments) (**h**) and in young animals without coelomocytes (*n* = 3 independent experiments) (**i**). **j**, LBP-2::tagRFP levels detected by western blot, with fold changes shown relative to corresponding levels at day 2 (*n* = 2 independent experiments). **k**, Quantification of LBP-2::tagRFP aggregation with RNAi targeting top 13 ECR candidates (25° C, day 6, *n* = 1 independent experiment). 'Ctrl (−)' denotes empty vector negative control; 'ctrl (+)' denotes *rab-5* RNAi positive control. **l**, Quantification of LBP-2::tagRFP aggregation in *tag-196*(*ok822*), *lys-3*(*tm2505*) and *dod-21*(*ok1569*) mutants (day 6, left graph) and in the *clec-1*(*tm1291*) mutant (day 2 as mutation was lethal afterwards, right graph) (*n* = 3 independent experiments). *P* values were determined by chi-square test (**h**, **l** left graph), two-sided Fisher's exact test (**i**, **l** right graph) and ordinal logistic regression (**k**). For blot source image, see Supplementary Fig. 1.

levels of soluble LBP-2 decreased with age (Fig. 1j, Extended Data Fig. 1g). Together, these observations show that secreted LBP-2 aggregates with age in the extracellular space.

To identify factors that are responsible for regulating protein aggregation in the extracellular space, we performed an RNA interference (RNAi) screen targeting genes encoding the secreted proteome. After retesting, 57 genes knocked-down by RNAi were confirmed to accelerate LBP-2 aggregation (Extended Data Fig. 2a, Supplementary Table 1). Secreted candidates identified in this screen are likely to act in the extracellular space to modulate LBP-2 aggregation rather than during its secretion, as knockdown of several endoplasmic reticulum (ER) proteostasis components had no effect. Using Phyre2, we predicted secondary structure and found 36 proteins with domains related to sequences of known structure including several with enzymatic activity (Supplementary Table 1). To validate the ECRs further, we focused on 13 candidates with the strongest effect (Fig. 1k, Supplementary Table 1). Aggregation of LBP-2 was accelerated when subjecting animals to RNAi targeting candidates during adulthood (Fig. 1k) or during development (Extended Data Fig. 2b, c), without affecting total LBP-2 levels (Extended Data Fig. 2d, e). Loss-of-function mutants of four candidates confirmed our RNAi-based observations (Fig. 1l). Knockdown of ECRs by RNAi did not significantly impair coelomocyte uptake of secreted GFP compared to positive control *dyn-1* knockdown[12] (Extended Data Fig. 3a), which demonstrates that increased protein aggregation is probably not due to defective endocytosis. Together, the RNAi screening approach targeting secreted proteins represents a valuable method to discover new components of the extracellular proteostasis network.

Next, we tested whether ECRs modify the aggregation of another secreted aggregation-prone protein LYS-7[13] with a predicted antimicrobial lysozyme function[15], which forms puncta with age in the pseudocoelom (Extended Data Fig. 3b, c). Ten out of thirteen ECRs targeted by RNAi repeatedly accelerated LYS-7 aggregation (Extended

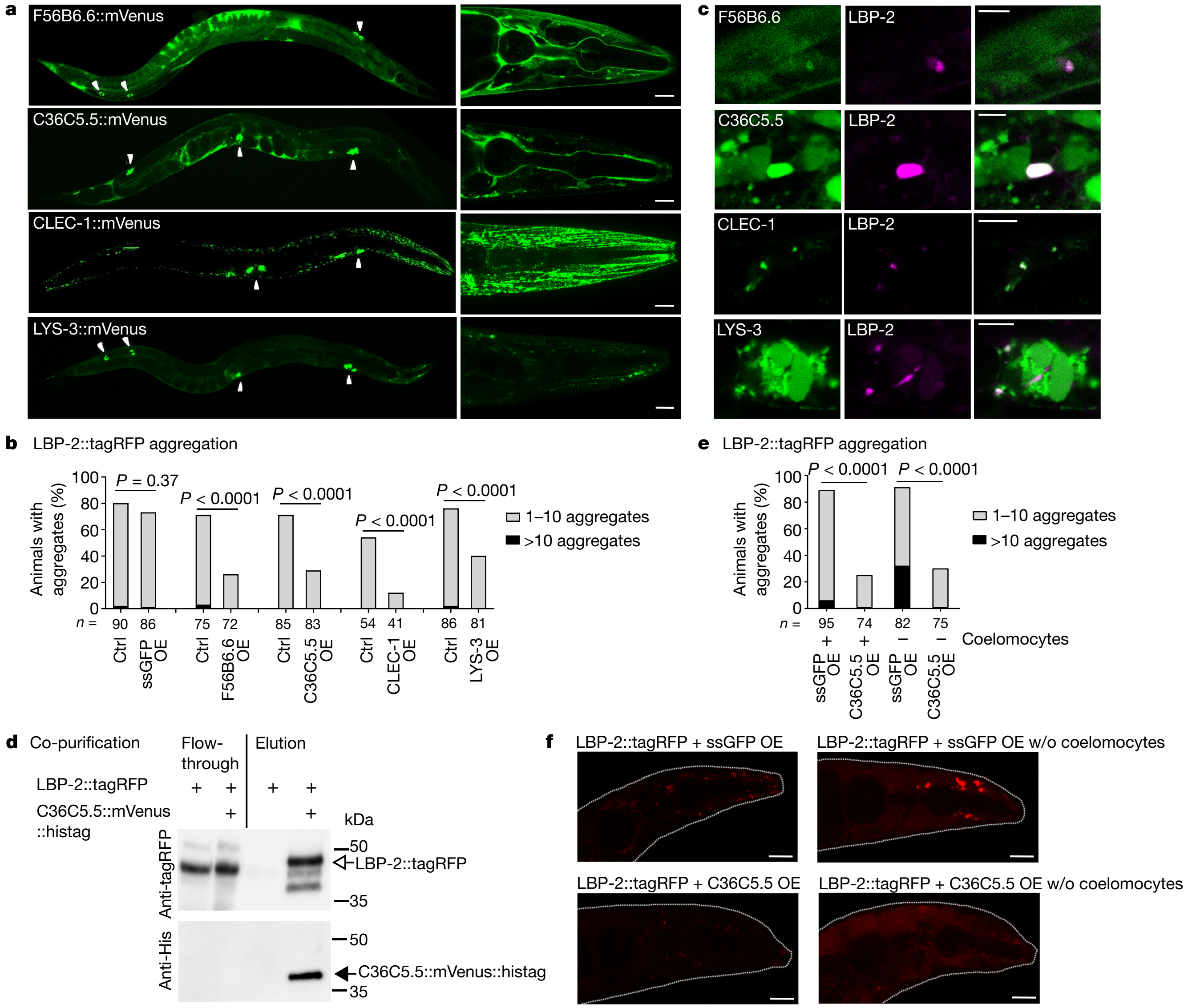


Fig. 2 | Overexpression of extracellular regulators prevents LBP-2 aggregation. a, Localization of mVenus-labelled ECRs, F56B6.6 (left $n = 6$, right $n = 12$ worms), C36C5.5 (left $n = 6$, right $n = 13$), CLEC-1 (left $n = 7$, right $n = 6$) and LYS-3 (left $n = 12$, right $n = 11$) in young animals. Arrowheads denote coelomocytes in whole animal (left). Maximum projection of head region (right). **b**, Quantification of LBP-2::tagRFP aggregation in animals overexpressing ECRs at day 6 ($n = 2$ independent experiments). 'Ctrl' indicates non-overexpressing siblings of ECR transgenics. **c**, Colocalization of LBP-2::tagRFP (magenta) and mVenus-tagged (green) F56B6.6 ($n = 11$ worms), C36C5.5 ($n = 7$), CLEC-1 ($n = 6$) or LYS-3 ($n = 10$) at day 3, 25 °C, to accelerate LBP-2 aggregation (single plane). **d**, Pull-down of C36C5.5::mVenus::histag co-purifies with LBP-2::tagRFP ($n = 2$ independent experiments). **e**, ECR C36C5.5 efficiently prevents LBP-2 aggregation at day 6 without coelomocytes ($n = 2$ independent experiments). **f**, ECR C36C5.5 maintains excessive LBP-2 diffuse in absence of coelomocytes (day 8; LBP-2 + secreted (ss) GFP overexpression (OE) $n = 17$ worms, LBP-2 + ssGFP OE without (w/o) coelomocytes $n = 22$, LBP-2 + C36C5.5 OE $n = 24$, LBP-2 + C36C5.5 OE w/o coelomocytes $n = 17$). Maximum projection of head region. Laser intensity 10%. Scale bars, 15 μm (**a**), 5 μm (**c**) or 20 μm (**f**). *P* values determined by two-sided Fisher's exact test (**b**) and chi-square test (**e**). For blot source image, see Supplementary Fig. 1.

Data Fig. 3d, Supplementary Table 1). Modulation of LYS-7 aggregation by CLEC-1 was shown by overexpression (Extended Data Fig. 3e). Thus, the ECRs identified in the present screen are likely to be constitutively active against a broad range of proteins aggregating in the extracellular space.

To assess whether overexpression of ECRs improves extracellular proteostasis, we randomly chose to overexpress in the body-wall muscle C36C5.5 and F56B6.6—two uncharacterized proteins with cysteine-rich sequences—and the C-type lectin CLEC-1. In addition, we overexpressed the lysozyme-like protein LYS-3. The ECRs tagged with mVenus were secreted and taken up by coelomocytes (Fig. 2a). C36C5.5 and F56B6.6 were mainly diffusely localized in the extracellular space, whereas CLEC-1 and LYS-3 had a more punctate pattern (Fig. 2a). Notably, overexpression of all four candidates significantly prevented LBP-2 aggregation, whereas secreted GFP alone had no effect (Fig. 2b, Extended Data Fig. 4a). To understand how ECRs modulate extracellular protein aggregation, we assessed whether they interact with LBP-2. Indeed, we observed co-localization of all four ECRs with LBP-2 aggregates (Fig. 2c), indicating a specific interaction as secreted GFP did not accumulate in LBP-2 aggregates (Extended Data Fig. 4b). Examination of the interaction between C36C5.5 and LBP-2 showed that LBP-2 efficiently co-purifies with C36C5.5 (Fig. 2d, Extended Data Fig. 4c). C36C5.5 maintained excessive LBP-2 diffusely distributed in the pseudocoelom of animals that lack coelomocytes (Fig. 2e, f). Thus, the previously uncharacterized C36C5.5 acts as a holdase chaperone by directly binding to and stabilizing aggregation-prone proteins and,

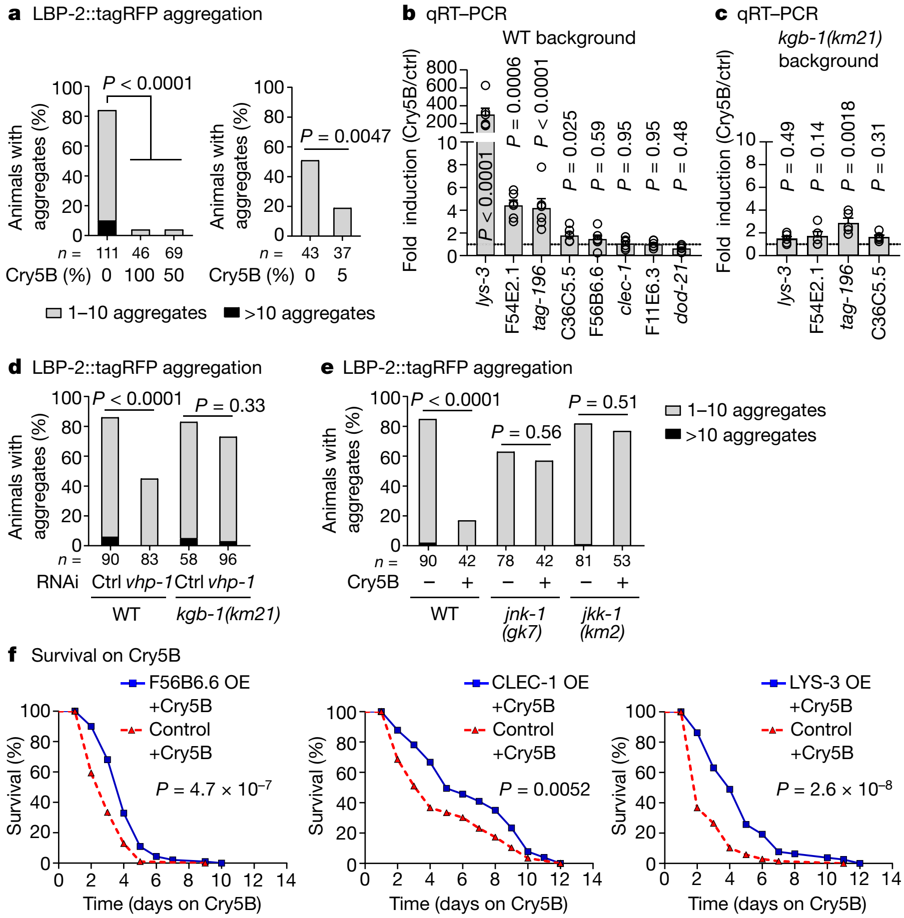


Fig. 3 | Pathogenic attack stimulates extracellular proteostasis through stress-activated MAP kinase signalling. a, Quantification of LBP-2::tagRFP aggregation during exposure to the Cry5B toxin (day 4, left graph, and day 5, right graph) ($n = 2$ independent experiments). **b**, **c**, Expression levels of ECRs, after 3 h exposure to 100% Cry5B, determined by qRT–PCR in the wild-type background ($n = 6$ biologically independent samples) and in the *kgb-1*(*km21*) loss-of-function mutant background ($n = 5$ biologically independent samples). Data are means ± s.e.m. **d**, Quantification of LBP-2::tagRFP aggregation at day 6 in wild-type and *kgb-1*(*km21*) mutant worms treated with empty vector or *vhp-1* RNAi. **e**, Quantification of LBP-2::tagRFP aggregation at day 6 in wild-type and *jnk-1*(*gk7*) or *jkk-1*(*km2*) loss-of-function mutants exposed to 5% Cry5B toxin. In **d** and **e**, $n = 2$ independent experiments. **f**, Survival analysis of F56B6.6-, CLEC-1- and LYS-3-overexpressing animals compared to non-overexpressing siblings subjected to 50% Cry5B exposure ($n = 2$ independent experiments, for detailed values see Extended Data Table 2). *P* values determined by chi-square test (**a**, left, and **d**), two-sided Fisher's exact test (**a**, right, and **e**), two-sided unpaired *t*-test with Welch's correction (**b**, **c**) or log-rank test (**f**).

together with the other overexpressed ECRs, stops extracellular protein aggregation.

We next investigated whether an improvement of extracellular proteostasis could have a beneficial effect on lifespan. Notably, the overexpression of F56B6.6 and LYS-3 extended *C. elegans* lifespan, whereas animals that lacked coelomocytes were shorter lived (Extended Data Fig. 5a, b, Extended Data Table 1). LBP-2 or secreted GFP overexpression alone had no effect on lifespan (Extended Data Fig. 5c). ECRs expression was differently regulated during ageing (Extended Data Fig. 5d) and notably *dod-21* (−24 fold) and F11E6.3 (−57 fold) were strongly downregulated with age, probably contributing to the collapse in extracellular proteome integrity with age. These results show that secreted ECRs can delay ageing and may affect other life history traits.

Bioinformatic analysis revealed a highly significant over-representation of ECRs among genes upregulated in response to diverse pathogens (Supplementary Table 2). This raises the possibility that enhancing extracellular proteostasis is part of the host response to pathogens, similar to expanding intracellular proteostasis capacity[6,8] such as inducing the ER unfolded protein response[9,16]. To investigate this, we exposed animals that overexpress LBP-2 to the virulence factor Cry5B, which forms pores in the intestinal plasma membrane and induces an extensive innate immune response[17,18]. Intoxication with Cry5B led to a strong reduction in LBP-2 aggregation at different concentrations (Fig. 3a) without reducing the transcription of *lbp-2* (Extended Data Fig. 6a). Two other non-lethal pathogens, *Microbacterium nematophilum* (which targets the anal cuticle) and *Bacillus atrophaeus*, also caused suppression of LBP-2 aggregation (Extended Data Fig. 6b). We examined whether extracellular proteostasis is upregulated, and found that three hours of exposure to Cry5B was sufficient to increase the expression of four out of eight ECRs (Fig. 3b). The JNK-like MAP kinase KGB-1, which is essential for survival on Cry5B, controls more than half of the genes induced[18]. Accordingly, we found that KGB-1 regulates ECR induction, as *kgb-1* mutants induced only one out of four ECRs upregulated in response to Cry5B, without strongly affecting basal levels of expression (Fig. 3c, Extended Data Fig. 6c). By contrast, the inhibition of VHP-1, a MAP kinase phosphatase that negatively regulates both KGB-1 and PMK-1[19,20], induced three out of eight ECRs (Extended Data Fig. 6d) and strongly reduced LBP-2 aggregation in a KGB-1-dependent manner (Fig. 3d). We also examined the role of another stress-activated MAP kinase, JNK-1. Although JNK-1 is not essential for survival on Cry5B[21], *jnk-1*(*gk7*) mutants were significantly shorter lived when exposed to Cry5B (Extended Data Fig. 6e). Notably, during Cry5B intoxication, loss-of-function of *jnk-1* or the upstream regulator *jkk-1* restored the aggregation of LBP-2 (Fig. 3e).

Thus, stress-activated MAPK signalling controls the upregulation of extracellular proteostasis during a pathogenic attack.

Finally, we investigated whether enhanced extracellular proteostasis confers a survival advantage against pathogens. Animals overexpressing F56B6.6, CLEC-1 or LYS-3 were strongly protected against lethality induced by Cry5B, with a more than 30% increase in viability (Fig. 3f, Extended Data Table 2). By contrast, reduced expression caused hypersensitivity. Similarly, RNAi targeting *tag-196* and F54E2.1, two ECRs strongly induced by Cry5B, as well as coelomocyte ablation accelerated *C. elegans* demise on the toxin. LBP-2 or secreted GFP overexpression alone did not negatively influence survival on Cry5B (Extended Data Fig. 6f–i, Extended Data Tables 2, 3). To investigate how enhanced extracellular proteostasis helps the host to survive an initially localized pathogenic attack, we performed RNA sequencing on Cry5B-challenged and -unchallenged animals. First, we found that animals with enhanced extracellular proteostasis achieved a higher upregulation of pathogen-response genes during exposure to Cry5B, including genes such as collagens expressed in non-intestinal tissues. Second, there is no evidence that ECR upregulation itself triggers an immune response in basal conditions (Extended Data Fig. 7, Supplementary Table 3). These results suggest that a healthy extracellular proteome supports a transcriptional response to pathogens, in several tissues, without inducing it. Also, extracellular aggregates are probably a liability during the attack as increased extracellular protein aggregation is associated with hypersensitivity to Cry5B (Extended Data Fig. 6f, j). Optimizing extracellular proteostasis would sustain a systemic immune response by ensuring favourable pseudocoelomic conditions for intercellular signalling[22] and by keeping secreted immune factors functional. CLEC-1 secreted from body-wall muscles was highly effective in preventing aggregation of the immune factor LYS-7, which is mainly produced by the intestine[15] (Extended Data Fig. 3e).

We have uncovered the extracellular proteostasis network that regulates protein aggregation outside of the cell in *C. elegans*, probably through diverse mechanisms. Similar to the ER unfolded protein response[9,16,23], extracellular proteostasis has a role in ageing and in the response to pathogens. We propose that induction of extracellular proteostasis contributes to the host's systemic defence to counter pathogen propagation[24] from different sites of attack while alleviating proteotoxicity from enhanced immunity-related secretion. Notably, bioinformatic analysis revealed that half of the 57 ECRs have potential human orthologues of which the majority are expressed in the brain and have associations with neurodegenerative diseases, including some differentially regulated in cerebrospinal fluid from Alzheimer's disease[25] (Supplementary Table 4). This orthologue list is likely to be a valuable source for the discovery of human extracellular proteostasis components relevant for Alzheimer's disease. Strong evidence already supports the protective role of extracellular proteostasis in Alzheimer's disease, and both the extracellular chaperone clusterin and the microglia receptor TREM2, which act together to facilitate the removal of pathological amyloid-β, are risk factors for Alzheimer's disease[26–30]. Moreover, knowledge of the control over extracellular proteostasis in *C. elegans* can pave the way for the discovery of a similar master regulator in mammals.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2461-z.

1. Hoshino, A. et al. A novel function for proSAAS as an amyloid anti-aggregant in Alzheimer's disease. *J. Neurochem*. **128**, 419–430 (2014).
2. Helwig, M. et al. The neuroendocrine protein 7B2 suppresses the aggregation of neurodegenerative disease-related proteins. *J. Biol. Chem*. **288**, 1114–1124 (2013).
3. Wyatt, A. R., Yerbury, J. J., Ecroyd, H. & Wilson, M. R. Extracellular chaperones and proteostasis. *Annu. Rev. Biochem*. **82**, 295–322 (2013).
4. Genereux, J. C. et al. Unfolded protein response-induced ERdj3 secretion links ER stress to extracellular proteostasis. *EMBO J*. **34**, 4–19 (2015).
5. Tucker, H. M. et al. The plasmin system is induced by and degrades amyloid-beta aggregates. *J. Neurosci*. **20**, 3937–3946 (2000).
6. Ermolaeva, M. A. et al. DNA damage in germ cells induces an innate immune response that triggers systemic stress resistance. *Nature* **501**, 416–420 (2013).
7. Pellegrino, M. W. et al. Mitochondrial UPR-regulated innate immunity provides resistance to pathogen infection. *Nature* **516**, 414–417 (2014).
8. Reddy, K. C. *et al*. An intracellular pathogen response pathway promotes proteostasis in *C. elegans*. *Curr. Biol*. **27**, 3544–3553 (2017).
9. Richardson, C. E., Kooistra, T. & Kim, D. H. An essential role for XBP-1 in host protection against immune activation in *C. elegans*. *Nature* **463**, 1092–1095 (2010).
10. Klaips, C. L., Jayaraj, G. G. & Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *J. Cell Biol*. **217**, 51–63 (2018).
11. Altun, Z. F. & Hall, D. H. *Pericellular Structures* https://www.wormatlas.org/hermaphrodite/pericellular/Periframeset.html (2009).
12. Fares, H. & Greenwald, I. Genetic analysis of endocytosis in *Caenorhabditis elegans*: coelomocyte uptake defective mutants. *Genetics* **159**, 133–145 (2001).
13. David, D. C. et al. Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol*. **8**, e1000450 (2010).
14. Ruskamo, S. et al. Molecular mechanisms of Charcot-Marie-Tooth neuropathy linked to mutations in human myelin protein P2. *Sci. Rep*. **7**, 6510 (2017).
15. Evans, E. A., Kawli, T. & Tan, M. W. *Pseudomonas aeruginosa* suppresses host immunity by activating the DAF-2 insulin-like signaling pathway in *Caenorhabditis elegans*. *PLoS Pathog*. **4**, e1000175 (2008).
16. Bischof, L. J. et al. Activation of the unfolded protein response is required for defenses against bacterial pore-forming toxin in vivo. *PLoS Pathog*. **4**, e1000176 (2008).
17. Wei, J. Z. et al. *Bacillus thuringiensis* crystal proteins that target nematodes. *Proc. Natl Acad. Sci. USA* **100**, 2760–2765 (2003).
18. Kao, C. Y. et al. Global functional analyses of cellular responses to pore-forming toxins. *PLoS Pathog*. **7**, e1001314 (2011).
19. Mizuno, T. et al. The *Caenorhabditis elegans* MAPK phosphatase VHP-1 mediates a novel JNK-like signaling pathway in stress response. *EMBO J*. **23**, 2226–2234 (2004).
20. Kim, D. H. et al. Integration of *Caenorhabditis elegans* MAPK pathways mediating immunity and stress resistance by MEK-1 MAPK kinase and VHP-1 MAPK phosphatase. *Proc. Natl Acad. Sci. USA* **101**, 10990–10994 (2004).
21. Huffman, D. L. et al. Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. *Proc. Natl Acad. Sci. USA* **101**, 10995–11000 (2004).
22. Melo, J. A. & Ruvkun, G. Inactivation of conserved *C. elegans* genes engages pathogen- and xenobiotic-associated defenses. *Cell* **149**, 452–466 (2012).
23. Taylor, R. C. & Dillin, A. XBP-1 is a cell-nonautonomous regulator of stress resistance and longevity. *Cell* **153**, 1435–1447 (2013).
24. Wan, L. et al. Bacillus thuringiensis targets the host intestinal epithelial junctions for successful infection of *Caenorhabditis elegans*. *Environ. Microbiol*. **21**, 1086–1098 (2019).
25. Khoonsari, P. E. et al. Analysis of the cerebrospinal fluid proteome in Alzheimer's disease. *PLoS One* **11**, e0150672 (2016).
26. Lambert, J. C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet*. **41**, 1094–1099 (2009).
27. Harold, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet*. **41**, 1088–1093 (2009).
28. Yeh, F. L., Wang, Y., Tom, I., Gonzalez, L. C. & Sheng, M. TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of amyloid-beta by microglia. *Neuron* **91**, 328–340 (2016).
29. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med*. **368**, 107–116 (2013).
30. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. *N. Engl. J. Med*. **368**, 117–127 (2013).

Methods

Strains

Wild-type animals were *C. elegans* variety Bristol, N2. The transgenic strains and alleles used in this work are described in Supplementary Table 5. Genetic crosses were made to transfer transgenes to the appropriate genetic background. The presence of the mutant allele was verified by polymerase chain reaction (PCR). Nematodes were grown and handled following standard procedures, under uncrowded conditions, at 15 °C, on NGM (Nematode Growth Medium) agar plates seeded with *Escherichia coli* strain OP50. Eggs or parents to lay eggs were transferred to 20 °C or 25 °C at the start of the experiment. Experiments were conducted with hermaphrodites at 20 °C (or at 25 °C to obtain sterile DCD130, CF512 animals or to accelerate LBP-2 aggregation, as indicated in the figure legends). Day 1 of adulthood is defined as 24 h after the last larval stage L4.

Cloning and strain generation

Cloning was carried out using the Gateway system (Life Technologies). The *myo3* promoter was provided by B. Lee. The *lbp-2* and *lys-7* promoters were amplified from N2 total DNA extract. *lbp-2*, *lys-7* and *lys-3* genes were amplified from cDNA from N2. C36C5.5, F56B6.6 and *clec-1* genes were amplified from N2 total DNA extract. All constructs contain the *unc-54* 3′ UTR. The tagRFP vector was obtained from Evrogen (AXXORA). mVenus was generated by targeted mutation of the YFP gene and C-terminal HisTag (RGSH6) was added. Constructs were sequenced at each step. plbp-2::lbp-2::tagRFP, pmyo-3::F56B6.6::mVenus::hisTag, pmyo-3::C36C5.5::mVenus::hisTag, pmyo-3::clec-1::mVenus::hisTag, pmyo-3::lys-3::mVenus::hisTag were injected at 100 ng μl^{-1} (whole plasmid injected). plys-7::lys-7::tagRFP was injected at 50 ng μl^{-1} (whole plasmid) with the co-injection marker podr-1::CFP at 30 ng μl^{-1}. Constructs were injected into N2 animals, except C36C5.5, F56B6.6, *clec-1* and *lys-3* injected in DCD23 animals. plbp-2::lbp-2::tagRFP transgene was integrated using UV irradiation and was backcrossed four times into the wild-type N2 strain. Animals overexpressing pmyo-3::C36C5.5::mVenus::hisTag incorporated the transgene into their genome during culture.

Aggregation quantification in vivo

Starting with a population of approximately 100 synchronized worms, aggregation levels were determined between days 2 and 12 as reported in the figure legends using a Leica fluorescence microscope M165 FC with a Planapo 2.0× objective. The number of fluorescent-labelled LBP-2 or LYS-7 puncta was manually counted only in the head region (or tail region as indicated in figure legend) of the worms and animals were classified into three categories: animals with no puncta, animals with up to ten and more than ten.

RNAi screen

Using SignalP v.4.0 (http://www.cbs.dtu.dk/services/SignalP/)[31], we selected all genes which contain predicted signal peptides. To exclude membrane proteins that remain in the ER–Golgi or that are localized to the plasma membrane, we removed all genes with predicted transmembrane domains using the TMHMM bioinformatics resource (http://www.cbs.dtu.dk/services/TMHMM)[32]. RNAi clones matching these criteria and tested for changes in LBP-2 aggregation are found in Supplementary Table 6. All RNAi clones were obtained from the Marc Vidal RNAi feeding library or the Julie Ahringer RNAi feeding library (Source BioScience) and RNAi clones for top 13 ECRs were sequenced. The empty vector L4440 was used as control. RNAi by feeding was performed as previously described[33]. To collect large numbers of animals for the screen, 200,000 synchronized DCD130 transgenic worms were cultured at 25 °C in liquid culture with OP50-1 resistant against streptomycin until L4 stage as described previously[34]. After reaching the L4 stage, animals were repeatedly washed with sterile M9 to remove bacteria and 30 worms were placed on each RNAi NGM plate (triplicate plates for each RNAi clone). Plates were kept at 25 °C and the number of puncta was counted at day 6 of adulthood. RNAi clones were considered positive, when two out of three plates displayed higher numbers of fluorescent puncta in the head region of the worms compared to negative control (empty vector L4440). 162 positive clones were found in total. For the validation screen, hits from the whole screen were retested in quadruplicate. The repeat was performed with the similar procedures as for the whole screen. Fifty-seven RNAi clones induced higher number of fluorescent puncta in the head region of the worms in two out of four plates when compared to negative control (empty vector L4440). Thirteen RNAi clones induced higher LBP-2 puncta formation in all plates (seven out of seven) of the whole and repeat screens at 25 °C and in repeats at 20 °C with animals treated with RNAi during development. Of note, knockdown of several ER resident proteins with signal peptide involved in the ER unfolded protein response including key chaperones (*hsp-3*, *hsp-4* and *dnj-7*) and protein disulfide isomerases (*pdi-1*, *pdi-6* and C14B9.2)[35] did not accelerate LBP-2 aggregation, perhaps because misfolded proteins would be removed by ER assisted degradation.

Evaluation of endocytosis

Endocytosis was determined by evaluating the uptake of secreted GFP by coelomocytes. Animals were grouped into two categories: animals with or without GFP-labelled coelomocytes. For this, GS1912 transgenic animals, synchronized at larval stage L1, were placed on RNAi plates at 25 °C. The number of animals with or without fluorescent-labelled coelomocytes was counted at day 1 of adulthood.

Imaging

For microscopy, worms were immobilized in levamisole 100 mM (Sigma-Aldrich) on 2% agar pads. Using a Zeiss Axio Observer Z1 microscope and software ZEN 2.6 (2.6.76.000000), whole worm micrographs were taken with a 10× objective (EC Plan-NEOFLUAR 10×/0.3) and coelomocyte micrograph with a 63× oil objective (Plan-Apochromat 63×/1.40). For confocal analysis with a Leica SP8 confocal and software Leica Application Suite X (3.5.2.18963), worms were examined either with 63× glycerol objective or 40× oil objective (HC PL APO CS2 63×/1.30 or 40×/1.30). The tagRFP was detected using 555 nm as excitation and an emission range from 565 to 650 nm, mVenus using 515 nm as excitation and an emission range from 521 to 551 nm, GFP using 488 nm as excitation and an emission range from 500 to 550 nm. To visualize the muscle structures for confocal analysis, worms were collected at day 4 and fixed in 4% paraformaldehyde (PFA) for 10 min at room temperature. Worms were stained with Phalloidin-iFluor488 conjugate (1:50, Biomol) to visualize F-actin as described[36]. Phalloidin was visualized by excitation at 488 nm and with an emission window between 506 and 551 nm. Representative confocal images are displayed as maximum z-stack projection or single planes as described in the legends.

Insoluble protein extraction and western blot

Culture and protein extraction of *C. elegans* were performed as previously described[34]. To test the LBP-2 protein level with ageing, DCD130 were grown as a synchronized population at 25 °C in liquid culture. Worms were collected at days 2 and 10 of adulthood and separated from bacteria and dead worms by sucrose separation. Before freezing in liquid nitrogen, worms were resuspended 1:1 in RAB buffer (0.1 M MES, 1 mM EGTA, 0.1 mM EDTA, 0.5 mM $MgSO_4$, 0.75 M NaCl, 0.02 M NaF, Roche Complete Inhibitors 2x). For total protein extraction, ground worms were homogenized in 8 M urea, 2% SDS, 50 mM DTT, 50 mM Tris pH 7.4 at room temperature. For soluble and insoluble protein extraction, ground worms were homogenized in RIPA buffer (50 mM Tris pH 8, 150 mM NaCl, 5 mM EDTA, 0.5% SDS, 0.5% SDO, 1% NP-40, 1 mM PMSF, Roche Complete Inhibitors 1x). The detergent-soluble supernatant was collected after centrifugation for 20 min at 18,400*g*. The detergent-insoluble pellet was washed once with RIPA buffer and then solubilized in 8 M urea, 2% SDS, 50 mM DTT, 50 mM Tris pH 7.4 at room temperature. To test the LBP-2 protein level in animals treated with RNAi targeting ECRs, synchronized L1 DCD130 worms were grown on RNAi plates at 25 °C until day 4 of adulthood. 100 worms for each RNAi treatment were collected in loading buffer with reducing agent (NuPAGE, Thermofisher) and incubated at 70 °C for 10 min. Total protein staining was performed on the transfer blot using REVERT Total Protein Stain Kit (Li-Cor 926-11010) with staining solution diluted 2:1 with 100% methanol. The anti-tagRFP antibody (1:1,000, Evrogen) was used to detect LBP-2::tagRFP protein with secondary anti-rabbit antibody (1:10,000, IRDye680RD, Li-Cor). Detection was performed with Image Studio software (v.4.0.21) on Li-Cor Odyssey CLx imager in the 700 nm channel.

Co-purification

Cultures of *C. elegans* were performed as previously described[34]. In brief, DCD130 and DCD362 were grown as a synchronized population at 25 °C in liquid culture until day 3. Worms were collected and washed once in M9 plus 0.01% Triton, twice with M9, once with PBS before being frozen in liquid nitrogen and ground in a mortar. One-hundred milligrams of frozen ground worms were dissolved in 200 µl ice cold DPBS (Sigma) supplemented with protease inhibitors (Serva; Mix M) and syringed 20 times. The lysate was centrifuged for 5 min at 800*g* at 4 °C. The liquid layer between lipid lid and pellet was collected and centrifuged for 5 min at 2,200*g* at 4 °C. Then, 40 µl of HisTrap HP Sepharose (GE Healthcare; 50% solution pre-washed three times with DPBS supplemented with protease inhibitors) was added to the supernatant and incubated for 2 h at 4 °C. Flow-through was collected at 800*g* at 4 °C and beads were washed once with 500 µl DPBS with protease inhibitors followed by further washing steps in 500 µl of 40 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.7 and 500 µl of 40 mM sodium phosphate, 300 mM NaCl, 70 mM imidazole, pH 7.7. After a final washing step with 500 µl of 40 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.7 the proteins bound to the beads were eluted with 100 µl of 40 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 7.7. The resulting protein solutions were analysed by western blot. The target proteins were specifically detected with antibodies directed against tagRFP (1:4,000, Evrogen) or the 6xHistidine tag (1:2000, anti-His H3 antibody, Santa Cruz), followed by respectively anti-rabbit-IgG-POD conjugate antibody (Sigma) and anti-mouse-IgG-POD conjugate antibody (Sigma). POD was visualized by spraying the membrane with WesternBright™ ECL-spray (Advansta) following the manufacturer`s instructions. The luminescence signal was detected using an ImageQuant Las4000 system (GE Healthcare, v.1.2).

Exposure to Cry5B and pathogenic bacteria

For all experiments related to Cry5B and pathogenic bacteria, day 1 adult animals were transferred to lawns expressing Cry5B or containing pathogenic bacteria and maintained on these lawns at 20 °C for the duration of the experiment. *E. coli* strain JM103 expressing pQE9-Cry5B and *E. coli* strain JM103 containing empty vector pQE9 were cultured in liquid LB medium containing carbenicillin and IPTG as described[37]. Each culture was diluted to 0.2 ± 0.05 OD by adding LB medium to generate a stock solution for further dilutions. For 100% Cry5B expressing lawn, 100 µl of JM103 expressing Cry5B were spread on NGM plates with carbenicillin and IPTG. To obtain diluted Cry5B expressing lawns, *E. coli* JM103 containing empty vector was added to *E. coli* JM103 expressing Cry5B and then spread on NGM plates with carbenicillin and IPTG. Plates were incubated overnight at 25 °C. For Cry5B, we tested three different dilutions: 100, 50 and 5% (Fig. 3a). All the tested concentrations strongly reduced the aggregation of LBP-2 and for subsequent experiments, we chose 5% to minimize death related to Cry5B (Fig. 3e). *Bacillus atrophaeus* strain ATCC9372 (DSMZ Leibniz Institute), *Microbacterium nematophilum* strain CBX102 (CGC) and control *E. coli* strain OP50 were cultured in liquid LB medium at 37 °C overnight. LB medium was added to adjust each culture to the final concentration of 1 ± 0.05 OD and 100 µl was spread on NGM plates.

Lifespan and survival assays

For the lifespan assay, worms were grown on OP50-seeded NGM plates at 20 °C and were scored every day for live animals, dead animals (no longer responding to body touch) and censored animals (crawled off plates, contaminated, ruptured and bag of worms). For survival analysis in the presence of Cry5B toxin, worms were grown on OP50-seeded NGM plates until day 1 at 20 °C and then transferred onto plates with 50% *E. coli* strain JM103 expressing Cry5B or *E. coli* strain JM103 containing empty vector, maintained at 20 °C. The 50% dilution was chosen as this concentration allowed us to assess the effect of enhanced extracellular proteostasis on Cry5B survival during chronic exposition. For survival analysis with Cry5B and RNAi targeting ECRs, worms were grown on plates seeded with RNAi clones until day 1 at 20 °C except for *clec-1* knockdown where worms were grown on OP50 seeded NGM plates until L4 and then transferred to plates seeded with *clec-1* RNAi clone at 20 °C. At day 1, worms were transferred onto plates with a mix of 25% *E. coli* strain JM103 expressing Cry5B and 75% *E. coli* strain expressing RNAi. Survival was scored daily for live animals, dead animals and censored animals. For survival assays, bagged animals were included as dead. The numbers of animals are reported in Extended Data Tables 1–3.

qRT–PCR

Approximately 1,000 or more synchronized worms were used for each qRT–PCR experiment. Worms were rinsed from plates with M9 and washed twice with M9. Total RNA was prepared from worms using QIAZOL lysis reagent (Qiagen) and further purified with RNeasy Plus Universal Mini kit (Qiagen). cDNA was prepared by reverse transcription (SuperScript III, Invitrogen) using oligo-dT primers (Qiagen). qRT–PCR was performed on an ABI 7000 Instrument using SYBR Green detection (Applied Biosystems). For Cry5B experiments, day 1 adults were exposed for 3 h at 20 °C to either 100% *E. coli* strain JM103 expressing Cry5B or *E. coli* strain JM103 containing empty vector, as previously described[18]. In the wild-type background, six independent experiments were performed for each treatment. In the *kgb-1* (*km21*) background, five independent experiments were performed for each treatment. To quantify LBP-2::tagRFP expression level on Cry5B, four independent experiments were performed for each treatment. To test the expression levels of ECRs with *vhp-1* RNAi, synchronized L1 larvae were cultured at 20 °C until day 1 on *vhp-1* RNAi or empty vector plates. For each condition, four independent experiments were performed. *eft-2* was used as the qRT–PCR normalization control[21]. To test the expression levels of ECRs with age, synchronized L1 larvae were cultured at 25 °C until days 2 or 8 of adulthood (sterile background *fer-15*(*b26*)*II; fem-1*(*hc17*)*IV*). *pgk-1* was used as the qRT–PCR normalization control. For each condition, four independent experiments were performed. Primers used are listed in Supplementary Table 7.

RNA-seq experiments

Approximately 250 worms were collected per sample. Day 1 adults were exposed for 24 h at 20 °C to either 50% *E. coli* strain JM103 expressing Cry5B or *E. coli* strain JM103 containing empty vector. Worms were collected in M9, washed twice with M9 and frozen in liquid nitrogen. Total RNA was isolated using Direct-zol RNA Mini-Prep Kit following the manufacturers' instructions (Zymo Research). RNA concentration was quantified using Qubit (Invitrogen Life Technologies) and Nanodrop (PEQLAB Biotechnologie GmbH) measurements. RNA-seq libraries were prepared using TruSeq RNA library preparation kit v2 (Illumina Inc.) according to the manufacturer's instructions from 1 µg of total RNA in each sample. Libraries were quantified using Qubit and Bioanalyzer measurements (Agilent Technologies) and normalized to 2.5 nM. Samples were sequenced as 150-bp paired-end reads on multiplexed lanes of an Illumina HiSeq3000 (Illumina Inc.) resulting in 18–47 million read pairs per sample.

Analysis of RNA-seq data

Raw RNA-seq reads were mapped to the *C. elegans* reference assembly (PRJNA13758, WormBase v.WS250) by TopHat2 (v.2.0.14, default options)[38]. Expression levels were estimated as fragments per kilobase transcript per million mapped reads (FPKM) for each sample individually by Cufflinks (v.2.2.1, default options)[39]. FPKM values were then transformed into z-scores using the scale function of R (negative values were converted into zeros and the maximum z-score was set to 10). Z-scores were then used to perform hierarchical clustering (R function heatmap with scale = 'none' option and using single linkage clustering algorithm). In addition, to identify differentially expressed genes, we compared the ECR overexpressing samples against their respective control sample with the help of the Cuffdiff program (v.2.2.1,–library-norm-method classic-fpkm)[39]. We defined genes to be significantly differentially expressed if they showed an FDR-corrected $P < 0.1$ and an absolute fold change >2 in at least two of the ECR overexpressing samples.

Bioinformatic analysis

To bridge the evolutionary gap between the nematode and human proteins, a bioinformatics pipeline was built which derives consensus sequences with iterative sequence alignment using HMMER v.3.2.1. This algorithm was demonstrated to be more sensitive than regular sequence alignments in detecting remote orthologues (www.hmmer.org). The workflow was written using Python 3.5. Protein sequences were retrieved using the UniProt API. BioPython was used for handling of alignments produced by jackhmmer. Alignments were visualized using 'JavaScript Sequence Alignment Viewer'.

Genes identified in the screen were mapped to their corresponding Uniprot entries. For each entry a consensus sequence profile was iteratively created using the jackhmmer algorithm[40]. Each consensus profile was then used to search for human proteins and genes which match this profile. The potential human homologues were ranked following three criteria. First the relative alignment length (0 to 1, 1 means the potential human homologue has the same length as the consensus profile). Second the gene expression in the brain (−1 if not expressed, 0 if unknown, 1 if expressed)[41]. Third the association with known neuronal or mental diseases (sum of all scores)[42]. Brain gene expression values were taken from the Allen Human Brain Atlas[41]. Association with known diseases was taken from DisGeNET[42]. Only diseases that are classified as 'diseases of mental health' or 'nervous system disease' in the Disease Ontology[43] were taken into consideration. For each gene, the sum of the weighted evidence score was used. Human orthologues of ECR candidates were selected with ≥25% alignment and a total score ≥ 1.2. For three candidates (F11E6.3, *endu-1* and *lron-13*), orthologues are listed from Wormbase (Supplementary Table 4). Possible protein structures, folds and functions were predicted using the webserver Phyre2 with default settings[44]. Folds with over 90% confidence are reported (Supplementary Table 1). Several hits have predicted enzymatic functions such as hydrolase or transferase activity, which could change post-translational modifications to inhibit aggregation.

Statistical analysis

No statistical method was used to predetermine sample size and animals were randomly distributed between conditions. Aggregate counting experiments were performed blinded except with phenotypes precluding blinding. For analysis of aggregation, two-sided Fisher's exact test (GraphPad Prism, v.7.04) was performed to analyse two aggregation categories (animals with no puncta versus animals <10 puncta) and chi-square test (GraphPad) for three categories (animals with no puncta versus animals <10 puncta versus animals >10 puncta). When one aggregation category had less than three animals in both conditions, we performed two-sided Fisher´s exact test combining two categories together (animals with no puncta versus animals <10 puncta plus animals >10 puncta). For multiple aggregation categories and to analyse the effect on protein aggregation of multiple RNAi treatments compared to control RNAi, we used an ordinal logistic regression model, which was performed using R (v.3.6.0) and its MASS package (v.7.3–51.4)[45]. For two aggregation categories with multiple comparisons, we controlled for the false discovery rate with the Benjamini–Hochberg correction. Lifespan and survival assays were not performed blind and analysis was carried out by log-rank test with Bonferroni correction for multiple comparisons using OASIS 2 (https://sbi.postech.ac.kr/oasis2/)[46]. For qRT–PCR experiments, unpaired two-sided Student's t-test with Welch's correction (GraphPad) was used on ΔC_t of transcript levels, treated versus untreated from biologically independent samples. Enrichment analysis of ECRs in Supplementary Tables 2 and 3 was performed with WormExp (http://wormexp.zoologie.uni-kiel.de/wormexp/)[47], category Microbes, one-sided Fisher's exact test with Bonferroni correction $P < 0.05$. All numerical values used for graphs and detailed statistical analysis can be found in the Source data.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All relevant data are available and/or included with the manuscript or its Supplementary Information. RNA-sequencing data have been uploaded to the European Nucleotide Archive under the study accession PRJEB36386. Source data are provided with this paper.

Code availability

The source code for the bioinformatics analysis of homologues is available at https://github.com/Ashafix/C_Elegans_Homologs.

31. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
32. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol*. **305**, 567–580 (2001).
33. Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS Genet*. **1**, 119–128 (2005).
34. Groh, N. et al. Methods to study changes in inherent protein aggregation with age in *Caenorhabditis elegans*. *J. Vis. Exp*. **129**, 56464 (2017).
35. Shen, X., Ellis, R. E., Sakaki, K. & Kaufman, R. J. Genetic interactions due to constitutive and inducible gene regulation mediated by the unfolded protein response in *C. elegans*. *PLoS Genet*. **1**, e37 (2005).
36. Huang, C. et al. Intrinsically aggregation-prone proteins form amyloid-like aggregates and contribute to tissue aging in *Caenorhabditis elegans*. *eLife* **8**, e43059 (2019).
37. Bischof, L. J., Huffman, D. L. & Aroian, R. V. Assays for toxicity studies in *C. elegans* with Bt crystal proteins. *Methods Mol. Biol*. **351**, 139–154 (2006).
38. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. **14**, R36 (2013).
39. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol*. **28**, 511–515 (2010).
40. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
41. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
42. Piñero, J. et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).
43. Kibbe, W. A. et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res*. **43**, D1071–D1078 (2015).
44. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*. **10**, 845–858 (2015).
45. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer-Verlag New York, 2002).
46. Han, S. K. et al. OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research. *Oncotarget* **7**, 56147–56152 (2016).

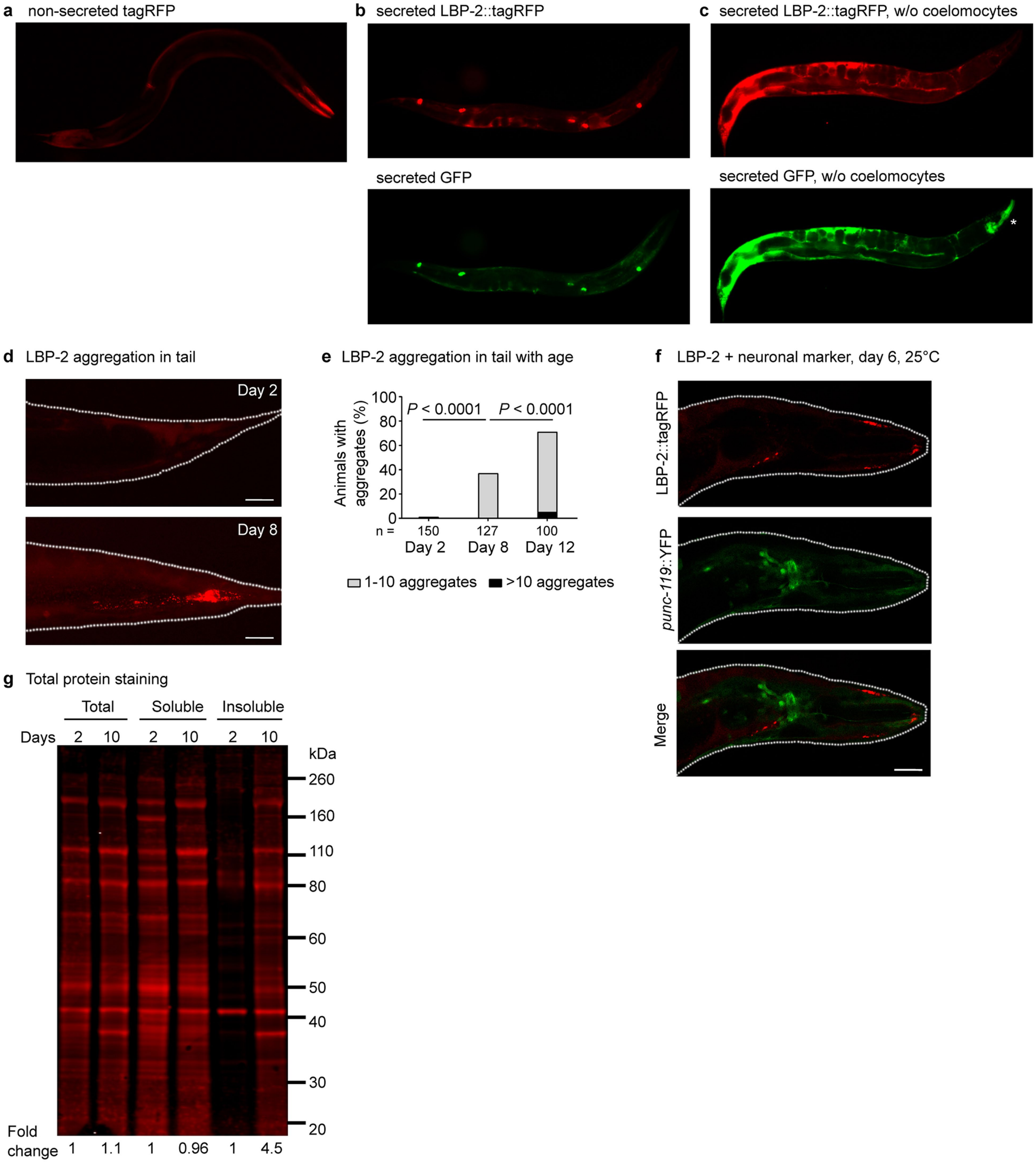
47. Yang, W., Dierking, K. & Schulenburg, H. WormExp: a web-based application for a *Caenorhabditis elegans*-specific gene expression enrichment analysis. *Bioinformatics* **32**, 943–945 (2016).

Acknowledgements We thank M. Schölling for help with statistics based on the ordinal logistic regression model. We thank C. Kenyon for providing some *C. elegans* strains and J. Fares for providing NP717. JM103 *E. coli* strains were provided by R. V. Aroian. Some *C. elegans* strains and *Microbacterium nematophilum* were provided by the *C. elegans* Genetics Center, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), the International *C. elegans* Gene Knockout Consortium and the National BioResource Project (NBRP). This work was supported by funding from the DZNE (D.C.D.), Max Planck Society (R.J.S.), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (SFB 1035 to M. H., DA 1906/4-1 to D.C.D.) and a Marie Curie International Reintegration Grant (322120 to D.C.D.).

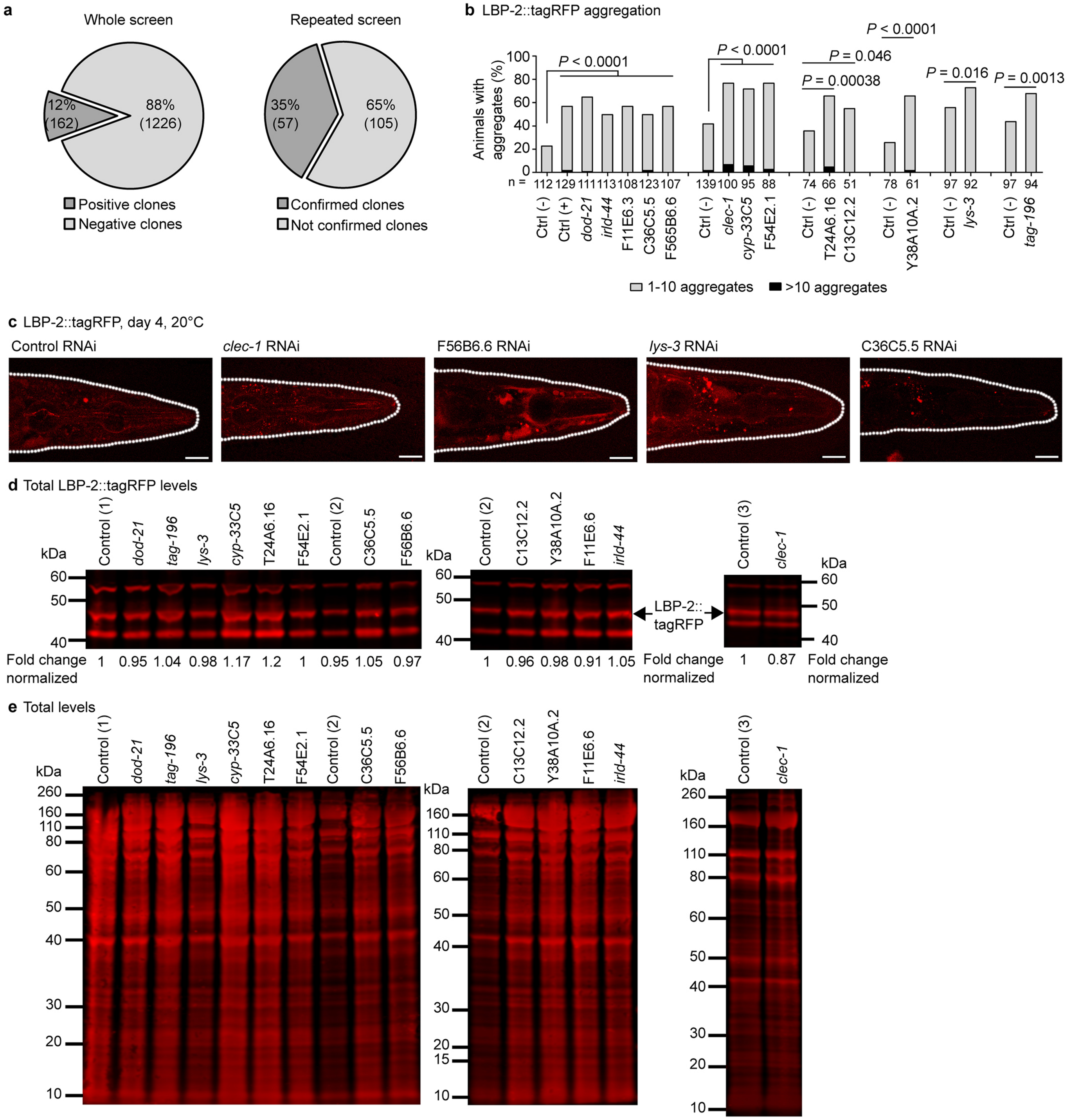
Author contributions I.G., A.S, M.H., C.H. and D.C.D. designed the experiments and interpreted data. I.G., A.S., R.J. and D.C.D. performed cloning and transgenic animal generation. R.J., A.S. and D.C.D. performed insolubility analysis of LBP-2 and R.J. carried out analysis of total LBP-2 levels with RNAi treatment. A.S. performed RNAi screen, LBP-2 aggregate counting and endocytosis evaluation. I.G. conducted LBP-2 aggregate counting, lifespan and survival analysis. A.S and I.G. independently confirmed the effects of ECR knockdown by RNAi on LBP-2 aggregation. J.L.B. and I.G. carried out aggregate counting of LYS-7. I.G. and D.C.D. performed image acquisition and analysis. M.H. conducted co-purification and western blots to detect interaction of LBP-2 and C36C5.5. I.G. collected animals for RT–qPCR and S.A., R.J. and C.H. performed RT–qPCR. W.R. carried out RNA sequencing. C.R. performed RNA-sequencing analysis. R.J.S. provided advice on RNA sequencing. M.P. performed homology analysis and structure prediction. D.C.D. and I.G. wrote the manuscript with contributions from R.J.S. and input from the other authors.

Competing interests The authors declare no competing interests.

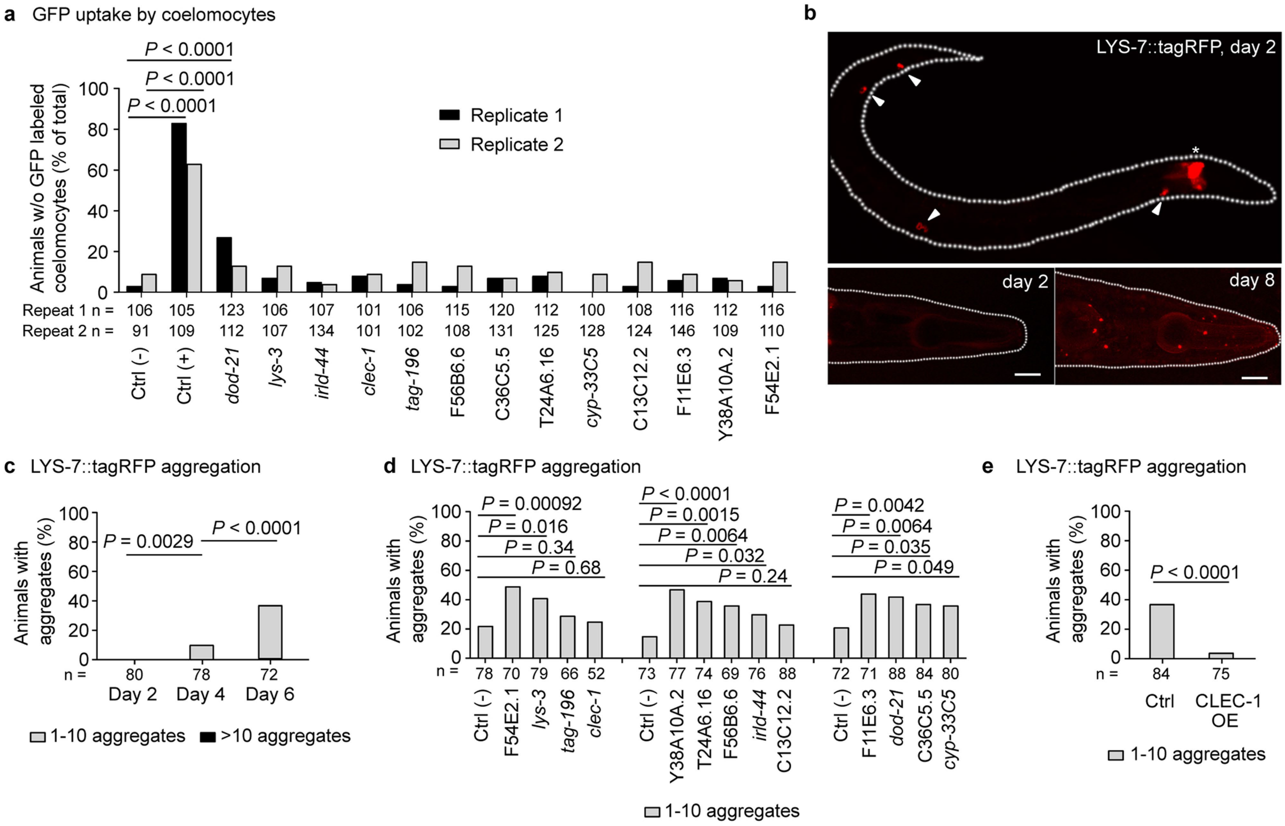
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2461-z.
Correspondence and requests for materials should be addressed to D.C.D.
Peer review information *Nature* thanks Carmen Nussbaum and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



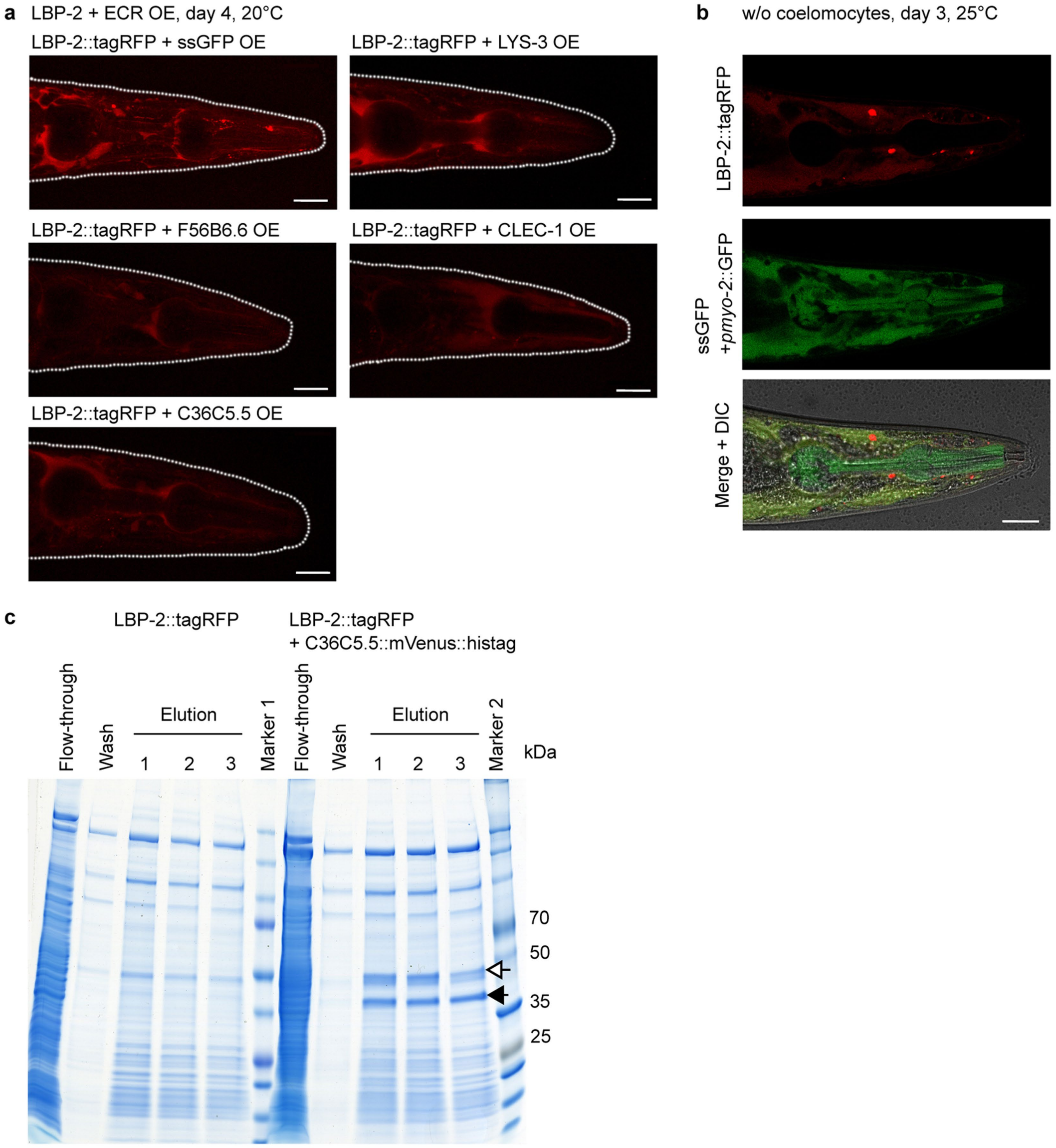
Extended Data Fig. 1 | Absence of coelomocytes causes LBP-2 to accumulate in pseudocoelom together with secreted GFP. **a**, LBP-2::tagRFP expression pattern in body-wall muscles (day 2, $n = 14$ worms). **b**, **c**, Secreted LBP-2::tagRFP and secreted GFP colocalize in day 2 animals with coelomocytes ($n = 16$ worms) (**b**) and accumulate in animals without coelomocytes ($n = 6$) (**c**). Asterisk indicates pharyngeal GFP reporter in animal without coelomocytes. **b**, **c**, Secreted GFP exposure, 5 times shorter in **c** versus **b**, secreted LBP-2::tagRFP, identical exposure. **d**, LBP-2::tagRFP puncta in tail region (day 2, $n = 15$ worms; day 8, $n = 14$). Maximum projection. Scale bar, 20 µm. **e**, Quantification of LBP-2::tagRFP aggregation with age in the tail ($n = 2$ independent experiments). *P* values determined by two-sided Fisher's exact test (day 8) and chi-square test (day 12). **f**, LBP-2 aggregates are separate from neurons ($n = 27$ worms). Scale bar, 20 µm. Single plane. **g**, Total protein stain of blot in Fig. 1j ($n = 2$ independent experiments) with fold changes quantified per fraction relative to levels in day 2. For blot source image, see Supplementary Fig. 1.



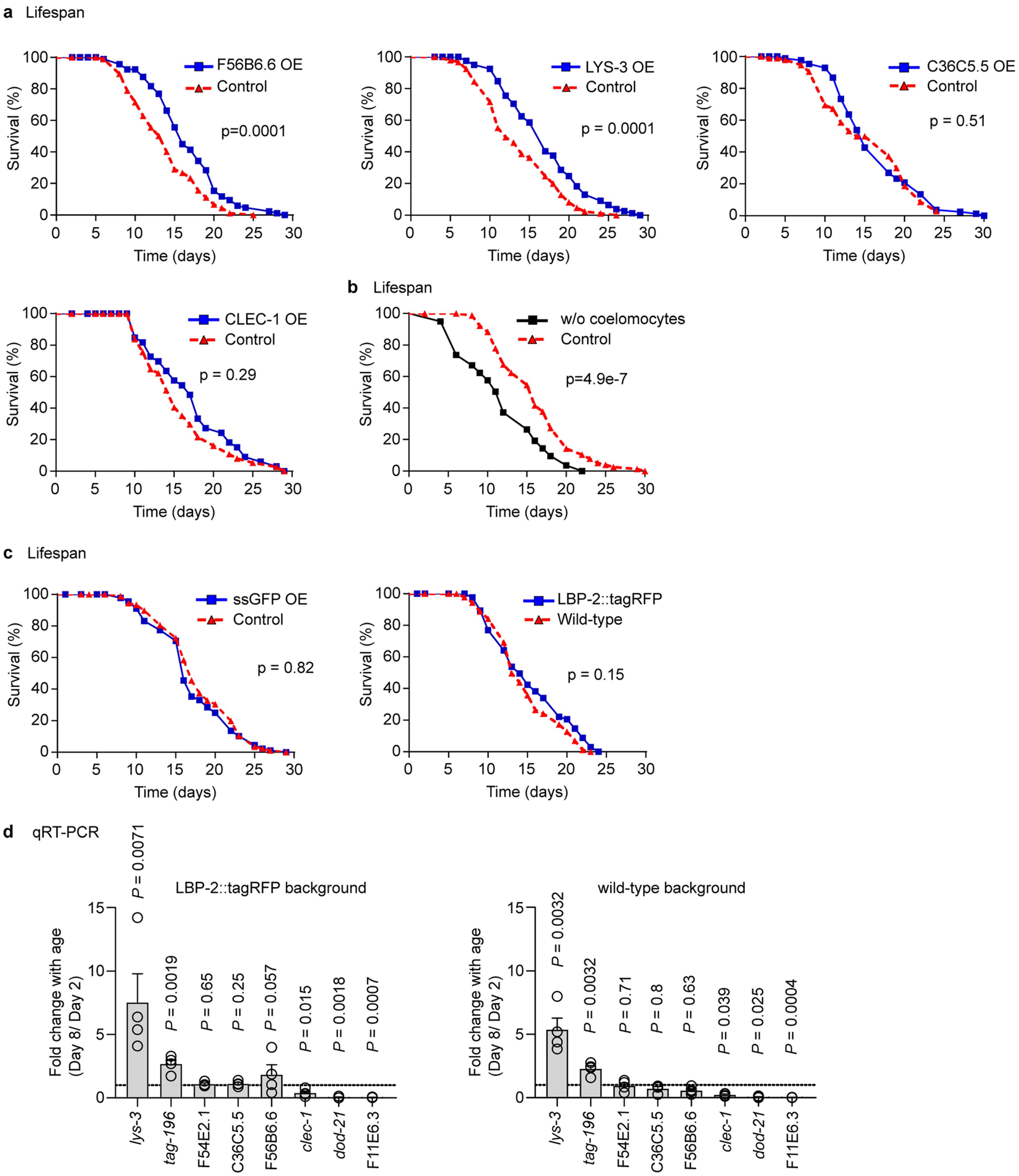
Extended Data Fig. 2 | RNAi targeting ECRs reproducibly enhances LBP-2 aggregation. a, Pie charts depict results from RNAi screen targeting genes encoding secreted factors and their effect on LBP-2 aggregation. **b**, Quantification of LBP-2::tagRFP aggregation with RNAi targeting top 13 candidates from egg in non-sterile background at day 4 ($n = 1$ independent experiment). Ctrl (−) is empty vector; ctrl (+) is *rme-1* RNAi. *P* values determined by ordinal logistic regression and for *lys-3* and *tag-196* RNAi treatment by two-sided Fisher's exact test. **c**, Maximum projection of head region of day 4 transgenics overexpressing LBP-2::tagRFP subjected to RNAi targeting a subset of ECRs (empty vector $n = 8$ worms, *clec-1* $n = 2$, F56B6.6 $n = 5$, *lys-3* $n = 7$, C36C5.5 $n = 5$). Scale bar, 20 µm. Laser intensity 8%. **d**, **e**, Downregulation of ECRs by RNAi does not change total levels of LBP-2::tagRFP ($n = 2$ independent experiments). Western blot detection of LBP-2::tagRFP in total fraction at day 4, 25 °C. Control 1, 2 and 3 are empty vector. Fold changes (in **d**) are normalized to total protein levels quantified by protein staining (**e**). For blot source images, see Supplementary Fig. 1.



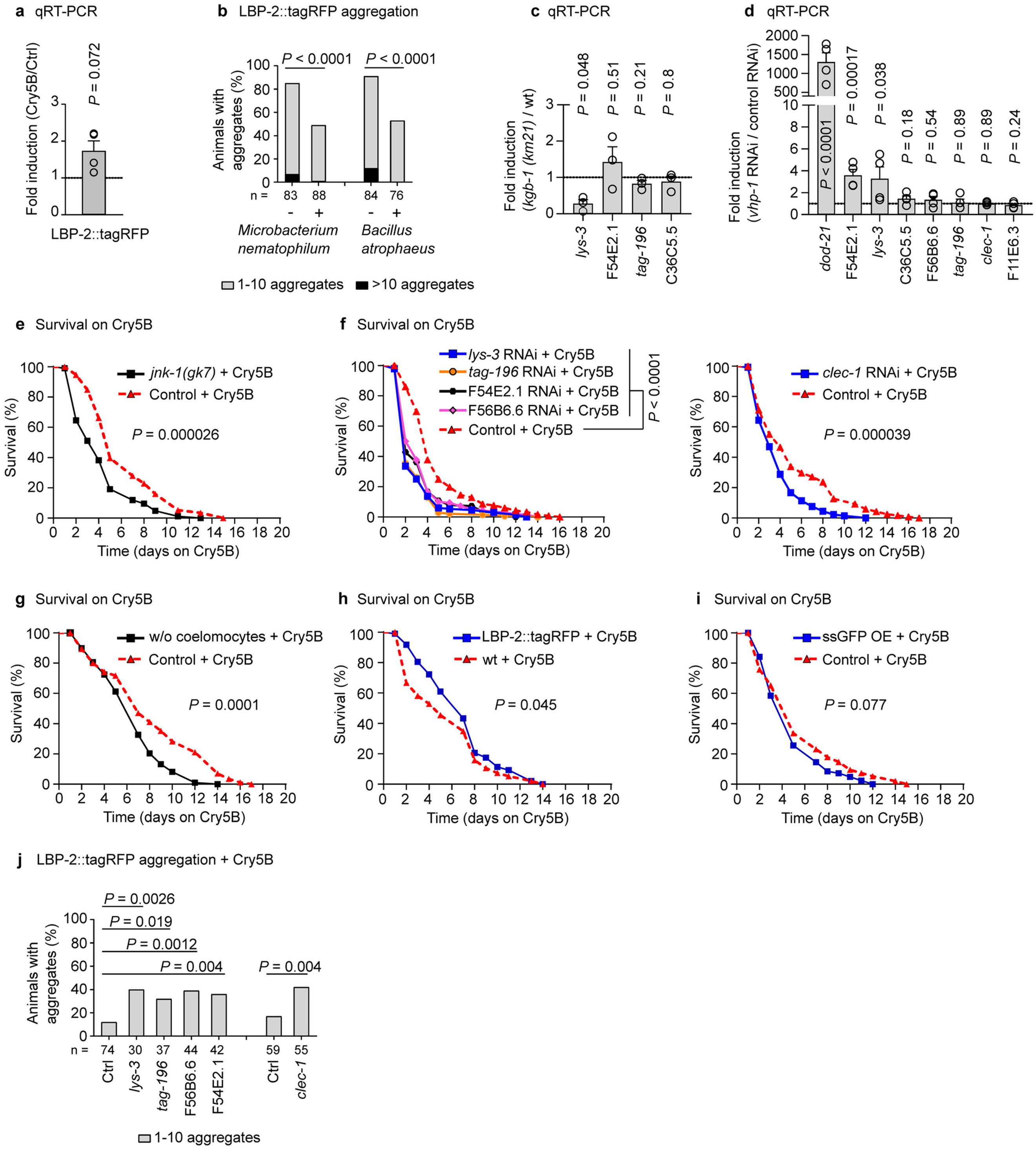
Extended Data Fig. 3 | ECRs regulate LYS-7 aggregation. a, Quantification of animals without GFP-labelled coelomocytes in transgenic animals expressing secreted GFP (GS1912) treated with RNAi targeting 13 top candidates ($n = 2$ independent experiments). Ctrl (−) denotes empty vector; ctrl (+) denotes *dyn-1* RNAi. **b**, LYS-7::tagRFP in young whole animal. Arrowheads indicate localization in coelomocytes, and asterisk indicates localization in anterior intestinal cells (top panel, $n = 10$ worms). LYS-7::tagRFP diffuse localization in head region of young animal (bottom left panel, $n = 19$) and puncta localized in head region of aged animal (bottom right panel, $n = 14$). Laser intensity 15%, maximum projection. Scale bar, 20 µm. **c**, Quantification of LYS-7::tagRFP aggregation with age ($n = 2$ independent experiments). **d**, Effect on LYS-7::tagRFP aggregation at day 6, 25 °C, with RNAi targeting top 13 candidates ($n = 4$ independent experiments). **e**, LYS-7 aggregation is reduced by CLEC-1 overexpression quantified at day 6 ($n = 2$ independent experiments). Ctrl indicates CLEC-1 non-overexpressing animals. P values determined by two-sided Fisher's exact test with Benjamini–Hochberg correction.



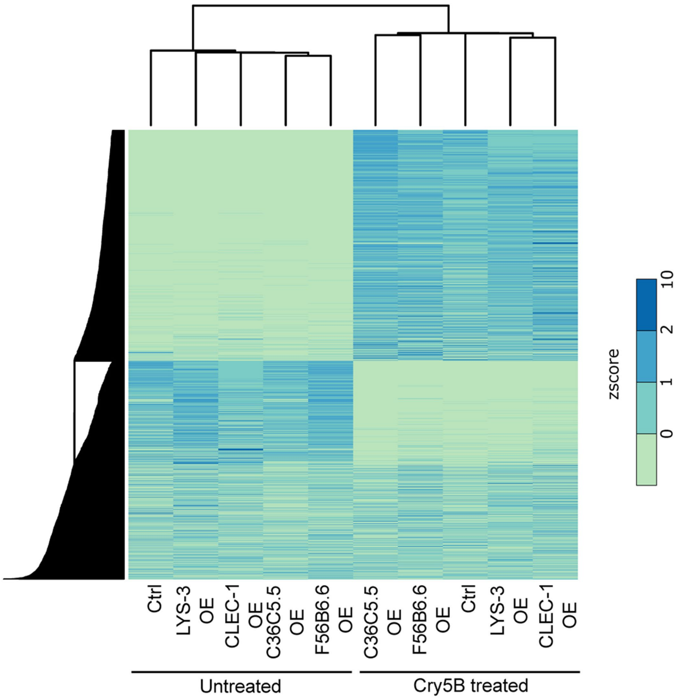
Extended Data Fig. 4 | ECR overexpression effectively prevents extracellular protein aggregation. **a**, Overexpression of ECRs reduces LBP-2 aggregation (ssGFP *n* = 13 worms, LYS-3 *n* = 27, F56B6.6 *n* = 21, CLEC-1 *n* = 19, C36C5.5 *n* = 24). Scale bar, 20 µm. Maximum projection, laser intensity 8%. **b**, Secreted GFP does not accumulate in LBP-2 aggregates (*n* = 21 worms). Scale bar, 20 µm. Single plane. **c**, Coomassie staining of co-purification of C36C5.5 with LBP-2 (*n* = 2 independent experiments). Open arrow indicates LBP-2::tagRFP; closed arrow denotes C36C5.5::mVenus::histag. Three independent co-purification experiments with the same starting material are shown (elution 1–3).



Extended Data Fig. 5 | Extracellular proteostasis influences ageing and is differentially regulated during ageing. a, F56B6.6-overexpressing and LYS-3-overexpressing animals are long-lived compared with non-overexpressing siblings (*n* = 2 independent experiments). **b**, Animals lacking coelomocytes are short-lived compared with control animals (*n* = 3 independent experiments). **c**, Secreted GFP (*n* = 2 biologically independent samples) and LBP-2 overexpression (*n* = 1 independent experiment) do not influence lifespan. *P* values were determined by log-rank test (**a**–**c**). For detailed values, see Extended Data Table 1. **d**, Changes in expression levels of ECR candidates with age (day 8 versus day 2 at 25 °C). Left, expression level in LBP-2 overexpressing sterile animals. Right, expression level in wild-type sterile animals. Data are means ± s.e.m. of *n* = 4 biologically independent samples. *P* values were determined by two-sided unpaired *t*-test with Welch's correction.



Extended Data Fig. 6 | Impairing extracellular proteostasis accelerates intoxication-related mortality. a, Expression level of LBP-2::tagRFP is not reduced after 3 h exposure to 100% Cry5B. Data are mean ± s.e.m. of $n = 4$ biologically independent samples. **b**, Quantification of LBP-2::tagRFP aggregation upon exposure to *Microbacterium nematophilum* and *Bacillus atrophaeus* at day 6 ($n = 2$ independent experiments). **c**, Expression level of four selected ECR candidates in unchallenged conditions with control pQE9 empty vector in *kgb-1*(*km21*) mutant versus wild-type background. **d**, Expression levels of eight selected ECRs with *vhp-1* versus control RNAi. Data are mean ± s.e.m. of $n = 3$ (**c**) and $n = 4$ (**d**) biologically independent samples. **e**, Survival analysis of LBP-2::tagRFP transgenics on *jnk-1*(*gk7*) versus wild-type background subjected to 50% Cry5B ($n = 2$ independent experiments). **f**, Survival analysis of LBP-2::tagRFP transgenics subjected to 25% Cry5B with RNAi targeting selected ECRs ($n = 2$ independent experiments). **g**, Survival analysis of secreted GFP transgenics with versus without coelomocytes, subjected to 50% Cry5B ($n = 4$ independent experiments). **h**, Survival analysis of LBP-2::tagRFP transgenics versus N2 wild-type subjected to 50% Cry5B ($n = 1$ independent experiment). **i**, Survival analysis of LBP-2::tagRFP transgenics with and without secreted GFP overexpression subjected to 50% Cry5B ($n = 2$ biologically independent samples). **j**, Increased LBP-2::tagRFP aggregation at day 4 of adulthood during exposure to 25% Cry5B and treatment with RNAi targeting selected ECRs compared to empty vector ($n = 2$ independent experiments). *P* values determined by two-sided unpaired *t*-test with Welch's correction (**a**, **c**, **d**), chi-square test (**b**) and two-sided Fisher's exact test with Benjamini–Hochberg correction (**j**), and log-rank test with Bonferroni correction (**e**–**i**). For detailed values see Extended Data Tables 2, 3.



Extended Data Fig. 7 | ECR overexpression does not induce a global immune response. Heat map of *z*-score-normalized expression data (rows correspond to genes, columns indicate samples). The dendrogram on the top shows that RNA-seq samples cluster by treatment (Cry5B versus control).

Extended Data Table 1 | Lifespan assays

Experiments	Strain name	N total / N censored	Mean lifespan ± s.e.m (days)	% change	P-value (log-rank)	Bonferroni P-value	Figure
# 1 Control *	DCD312	100 / 8	13.7 ± 0.44				Extended Data Figure 5a
# 1 F56B6.6 OE		100 /15	16.6 ± 0.5	+21.2%	p=0.0001		
# 2 Control *	DCD312	50 / 9	13.1 ± 0.7				
# 2 F56B6.6 OE		50 / 18	16.5 ± 0.96	+26%	p= 0.0086		
# 1 Control *	DCD401	100 / 14	13.67 ± 0.58				Extended Data Figure 5a
# 1 LYS-3 OE		100 / 23	16.9 ± 0.51	+23.6%	p=0.0001		
# 2 Control *	DCD401	100 / 10	14.09 ± 0.56				
# 2 LYS-3 OE		100 / 22	16.28 ± 0.58	+15.5%	p= 0.021		
# 1 Control *	DCD311	100 / 18	15.4 ± 0.57				Extended Data Figure 5a
# 1 C36C5.5 OE		100 / 18	16.1 ± 0.56	+4.5%	p= 0.51		
# 2 Control *	DCD311	50 / 22	14 ± 0.5				
# 2 C36C5.5 OE		50 / 15	15.2 ± 0.68	+8.6%	p= 0.13		
# 1 Control *	DCD308	50 / 13	15.7 ± 0.83				Extended Data Figure 5a
# 1 CLEC-1 OE		50 / 17	17.1 ± 0.94	+8.9%	p=0.29		
# 2 Control *	DCD308	50 / 17	16.6 ± 0.84				
# 2 CLEC-1 OE		50 / 12	15.7 ± 0.66	-5.4%	p=0.28		
# 1 Control	GS1912	100 / 23	16.03 ± 0.56				Extended Data Figure 5b
# 1 w/o coelomocytes	NP717	100 / 12	11.73 ± 0.52	-27%	p=4.9e-7		
# 2 Control	GS1912	100 / 18	14.93 ± 0.59				
# 2 w/o coelomocytes	NP717	100 / 8	11.35 ± 0.48	-24%	p=0.0000015		
# 3 Control	GS1912	100 / 10	14.17 ± 0.5				
# 3 w/o coelomocytes	NP717	100 / 4	11.57 ± 0.52	-18.3%	p=0.0022		
# 1 Control	DCD23	100 / 14	17.84 ± 0.5				
# 1 ssGFP OE (clone 2)	DCD371	100 / 12	17.12 ± 0.5	-4%	p=0.41	p=0.82	Extended Data Figure 5c
# 1 ssGFP OE (clone 1)	GS1912 / DCD23	100 / 10	16.63 ± 0.47	-6.8%	p=0.09	p=0.18	
# 1 wild-type	N2	100 / 13	14.6 ± 0.43				Extended Data Figure 5c
# 1 LBP-2::tagRFP	DCD23	100 / 26	15.23 ± 0.53	+4.3%	p=0.15		

Asterisk indicates siblings without the ECR overexpression.

Extended Data Table 2 | Survival assays on Cry5B

Experiments	Strain name	N total / N censored	Mean survival ± s.e.m (days on Cry5B)	% change	P-value (log-rank)	Bonferroni P-value	Figure
# 1 Control + Cry5B *	DCD312	100 / 6	3.09 ± 0.12				Figure 3f
# 1 F56B6.6 OE + Cry5B		100 / 9	4.12 ± 0.15	+33%	p=4.7e-7		
# 2 Control + Cry5B *	DCD312	100 / 4	4.61 ± 0.21				
# 2 F56B6.6 OE + Cry5B		100 / 14	6.24 ± 0.21	+35.4%	p=0.0000024		
# 1 Control + Cry5B *	DCD401	100 / 32	2.88 ± 0.19				Figure 3f
# 1 LYS-3 OE + Cry5B		100 / 22	4.7 ± 0.26	+63.2%	p=2.6e-8		
# 2 Control + Cry5B *	DCD401	70 / 10	4.74 ± 0.37				
# 2 LYS-3 OE + Cry5B		70 / 17	6.17 ± 0.36	+30.2%	p=0.012		
# 1 Control + Cry5B *	DCD308	75 / 18	4.75 ± 0.38				Figure 3f
# 1 CLEC-1 OE + Cry5B		75 / 31	6.39 ± 0.42	+34.5%	p= 0.0052		
# 2 Control + Cry5B *	DCD308	100 / 9	4.6 ± 0.24				
# 2 CLEC -1 OE + Cry5B		100 / 11	5.52 ± 0.23	+20%	p= 0.025		
# 1 Control + Cry5B *	DCD311	100 / 7	5.01 ± 0.26				
# 1 C36C5.5 OE + Cry5B		100 / 13	5.27 ± 0.23	+5.2%	p=0.7		
# 2 Control + Cry5B *	DCD311	75 /19	4.62 ± 0.43				
# 2 C36C5.5 OE + Cry5B		75 / 5	5.16 ± 0.38	+11.7%	p=0.41		
# 1 Control + Cry5B	DCD23	100 / 34	6.26 ± 0.38				Extended Data
# 1 *jnk-1(gk7)* + Cry5B	DCD360	100 / 11	4.25 ± 0.27	-31.1%	p=0.000026		Figure 6e
# 2 Control + Cry5B	DCD23	100 / 5	4.61 ± 0.28				
# 2 *jnk-1(gk7)* + Cry5B	DCD360	100 / 6	3.66 ± 0.2	-20.6%	p=0.006		
# 1 Control + Cry5B	GS1912	100 / 15	8.21 ± 0.46				Extended Data
# 1 w/o coelomocytes + Cry5B	NP717	100 / 2	6.51 ± 0.29	-20.7%	p=0.0001		Figure 6g
# 2 Control + Cry5B	GS1912	100 / 10	7.91 ± 0.34				
# 2 w/o coelomocytes + Cry5B	NP717	100 / 0	6.56 ± 0.27	-17.1%	p=0.0025		
# 3 Control + Cry5B	GS1912	100 / 7	7.52 ± 0.41				
# 3 w/o coelomocytes + Cry5B	NP717	100 / 0	6.27 ± 0.3	-16.6%	p=0.0061		
# 4 Control + Cry5B	DCD378	100 / 1	8.97 ± 0.42				
# 4 w/o coelomocytes + Cry5B	DCD379	100 / 0	5.91 ± 0.27	-34.1%	p=0		
# 1 wild-type + Cry5B	N2	100 / 5	5.5 ± 0.35				Extended Data
# 1 LBP-2::tagRFP + Cry5B	DCD23	100 / 3	6.79 ± 0.32	+23.5%	p=0.045		Figure 6h
# 1 Control + Cry5B	DCD23	100 / 5	5.59 ± 0.35				Extended Data
# 1 ssGFP OE (clone 2) + Cry5B	DCD371	100 / 7	4.86 ± 0.26	-13.1%	p=0.077	p=0.15	Figure 6i
# 1 ssGFP OE (clone 1) + Cry5B	GS1912 / DCD23	100 / 18	4.9 ± 0.28	-12.3%	p=0.092	p=0.18	

Asterisk indicates siblings without the ECR overexpression.

Extended Data Table 3 | Survival assays on Cry5B and ECR RNAi

Experiments	Strain name	N total / N censored	Mean survival ± s.e.m (days on Cry5B)	% change	Log-Rank test	Bonferroni P-value	Figure
# 1 L4440 + Cry5B	DCD23	144 / 25	4.99 ± 0.27				
# 1 F54E2.1 RNAi + Cry5B	DCD23	144 / 6	3.38 ± 0.2	-32.3%	p=2.2e-7	p=8.7e-7	
# 1 *tag-196* RNAi + Cry5B	DCD23	144 / 5	2.94 ± 0.15	-41.1%	p=0	p=0	Extended Data Figure 6f
# 1 *lys-3* RNAi + Cry5B	DCD23	144 / 5	3.04 ± 0.19	-39.1%	p=0	p=0	
# 1 F56B6.6 RNAi + Cry5B	DCD23	144 / 5	3.5 ± 0.21	-29.9%	p=0.0000013	p=0.0000051	
# 2 L4440 + Cry5B	DCD23	144 / 12	4.95 ± 0.28				
# 2 F54E2.1 RNAi + Cry5B	DCD23	144 / 7	2.99 ± 0.11	-39.6%	p=0	p=0	
# 2 *tag-196* RNAi + Cry5B	DCD23	144 / 8	2.65 ± 0.11	-46.5%	p=0	p=0	
# 2 *lys-3* RNAi + Cry5B	DCD23	144 / 9	3.3 ± 0.21	-33.3%	p=0.0000012	p=0.0000048	
# 2 F56B6.6 RNAi + Cry5B	DCD23	144 / 7	3.27 ± 0.16	-33.9%	p=1.6e-7	p=6.4e-7	
# 1 L4440 + Cry5B	DCD23	144 / 20	5.37 ± 0.33				
# 1 *clec-1* RNAi + Cry5B	DCD23	144 / 12	3.85 ± 0.19	-28.3%	p=0.000039		Extended Data Figure 6f
# 2 L4440 + Cry5B	DCD23	144 / 1	4.27 ± 0.29				
# 2 *clec-1* RNAi + Cry5B	DCD23	144 / 1	3.08 ± 0.17	-27.9%	p=0.0001		

nature research

Corresponding author(s): Della C. David

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Leica Application Suite X v. 3.5.2.18963, ZEN 2.6 v. 2.6.76.00000, ImageQuant Las4000 v.1.2, Image Studio v. 4.0.21
Data analysis	GraphPad software v. 7.04, R v. 3.6.0 and its MASS package v. 7.3-51.4, SignalP 4.0, TMHMM Server v. 2.0, HMMER v3.2.1, Phyre2, Python 3.5, Biopython v. 1.72, OASIS 2, Photoshop CS5 v. 12.1, ImageJ v. 1.49, https://github.com/Ashafix/C_Elegans_Homologs, TopHat2 version 2.0.14, Cufflinks version 2.2.1, Cuffdiff program version 2.2.1., WormExp http://wormexp.zoologie.uni-kiel.de/wormexp/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available within the paper and the Supplementary Information files. Raw reads for RNA sequencing data are available at the European Nucleotide Archive under the study accession PRJEB36386.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size. For each experiment the n values were reported in the legend, source data file and Extended Data Tables 1-3. We used standard sample sizes based on previous experiments and literature and that were needed to perform the statistical tests.
Data exclusions	In general, data were not excluded. For aggregation counting, all worms that died during the experiments, crawled off the plate, exploded or bagged were not included in the data (pre-established criteria). For lifespan assay, worms that crawled off the plate, exploded or bagged were not included in the data (pre-established criteria). For survival assay, worms that crawled off the plate or exploded were not included in the data (pre-established criteria). This exclusion criteria is extensively used in the field.
Replication	Experiments were repeated at least two times. All results were successfully and reliably repeated (see source data with repeats). Effects of top 13 ECRs were independently confirmed by two investigators.
Randomization	Worms were randomly distributed into treatment and control groups.
Blinding	Investigators were blinded to the genotype or RNAi treatments when feasible. For obvious mutant phenotypes or for strong overexpression of ECRs visible in the red channel of the fluorescence dissecting microscope, this was not possible. Survival and lifespan assays were not performed blind as death is an unambiguous read-out.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Unique biological materials
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Anti-tRFP, Evrogen, cat.# AB233, Lot. 23301301013, dilution 1:1000 and 1:4000. anti-His H3, Santa Cruz Biotechnology, cat.# sc-8036, Lot. 10414, dilution 1:2000. Secondary anti-rabbit IRDye680RD, LI-COR, cat.# 927-68071, Lot. C51104-08, dilution 1:10,000.
Validation	http://evrogen.com/products/antibodies/AB-tRFP.shtml, https://www.scbt.com/scbt/product/his-probe-antibody-h-3

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Caenorhabditis elegans hermaphrodites were the only experimental organism used in this study. RNAi screening was done at day 6 of adulthood. For aggregate counting, animals at day 2, day 4, day 6, day 8, day 12 of adulthood were used. Evaluation of endocytosis was performed at day 1 of adulthood. For insoluble protein extractions, animals of day 2 and day 10 were used. Co-purification was performed at day 3 of adulthood. For qRT-PCR, day 1 adults were used for all experiments except to test expression levels with age, where animals at day 2 and day 8 of adulthood were used. RNAseq was performed with day 1 adults.
Strains used are the following:
DCD1 uqEx1[plbp-2::lbp-2::tagrfp]
DCD23 uqIs5[plbp-2::lbp-2::tagrfp]
CF512 fer-15(b26) II; fem-1(hc17) IV
DCD130 fer-15(b26) II; fem-1(hc17) IV; uqIs5[plbp-2::lbp-2::tagrfp]
DCD320 uqEx55[plys-7::lys-7::tagrfp; podr-1::cfp]
GS1912 dpy-20(e1282) IV; aIrs37[pmyo-3::ssgfp; dpy-20(+)]
NP717 unc-119(ed3); arIs37[pmyo-3::ssgfp; dpy-20(+)]; cdIs32[pcc1::DT-A(E148D); unc-119(+); pmyo-2::gfp]
DCD378 arIs37[pmyo-3::ssgfp; dpy-20(+)]
DCD379 unc-119(ed3); arIs37[pmyo-3::ssgfp; dpy-20(+)]; cdIs32[pcc1::DT-A(E148D); unc-119(+); pmyo-2::gfp]
DCD308 uqEx52[pmyo-3::clec-1::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD311 uqEx53[pmyo-3::C36C5.5::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD312 uqEx54[pmyo-3::F56B6.6::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD363 uqIs27[pmyo-3::C36C5.5::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD362 fer15(b26) II; fem-1(hc17) IV; uqIs27[pmyo-3::C36C5.5::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD401 uqEx61[pmyo-3::lys-3::mVenus::histag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD381 unc-119(ed3); arIs37[pmyo-3::ssgfp; dpy-20(+)]; cdIs32[pcc1::DT-A(E148D); unc-119(+); pmyo-2::gfp]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD383 uqIs5[plbp-2::lbp-2::tagrfp]; arIs37[pmyo-3::ssgfp; dpy-20(+)]
DCD371 uqIs5[plbp-2::lbp-2::tagrfp]; arIs37[pmyo-3::ssgfp; dpy-20(+)]
DCD387 unc-119(ed3); cdIs32 [(pcc1::DT-A(E148D); unc-119(+); pmyo-2::gfp]; uqIs27[pmyo-3::C36C5.5::mVenus::hisTag]; uqIs5[plbp-2::lbp-2::tagrfp]
DCD403 uqEx62[plbp-2::tagrfp]
DCD398 tag-196(ok822) V; uqEx1[plbp-2::lbp-2::tagrfp]
DCD370 clec-1(tm1291) V; uqEx1[plbp-2::lbp-2::tagrfp]
DCD400 uqEx56[plys-7::lys-7::tagrfp; podr-1::cfp]; uqIs31[pmyo-3::clec-1::mVenus::histag]
DCD405 lys-3(tm2505) V; uqEx1[plbp-2::lbp-2::tagrfp]
DCD406 dod-21 & C32H11.1(ok1569) IV; uqEx1[plbp-2::lbp-2::tagrfp]
DCD359 jkk-1(km2) X; uqIs5[plbp-2::lbp-2::tagrfp]
DCD360 jnk-1(gk7) IV; uqIs5[plbp-2::lbp-2::tagrfp]
DCD361 kgb-1(km21) IV; uqIs5[plbp-2::lbp-2::tagrfp]
KU21 kgb-1(km21) IV
KU2 jkk-1(km2) X
VC8 jnk-1(gk7) IV
RB1384 dod-21 & C32H11.11(ok1569) IV
RB939 tag-196(ok822) V
FX02505 lys-3(tm2505) V
FX01291 clec-1(tm1291) V
DCD410 uqIs5[plbp-2::lbp-2::tagrfp]; uEx60 [punc-119::yfp; punc-119::sid-1]

Wild animals

The study did not involve wild animals

Field-collected samples

The study did not involve samples collected from the field

Fitness trade-offs incurred by ovary-to-gut steroid signalling in *Drosophila*

https://doi.org/10.1038/s41586-020-2462-y

Received: 7 September 2018

Accepted: 15 April 2020

Sara Mahmoud H. Ahmed[1,2,3], Julieta A. Maldera[1,2], Damir Krunic[1], Gabriela O. Paiva-Silva[4], Clothilde Pénalva[5], Aurelio A. Teleman[1,3]✉ & Bruce A. Edgar[1,2,5]✉

Sexual dimorphism arises from genetic differences between male and female cells, and from systemic hormonal differences[1–3]. How sex hormones affect non-reproductive organs is poorly understood, yet highly relevant to health given the sex-biased incidence of many diseases[4]. Here we report that steroid signalling in *Drosophila* from the ovaries to the gut promotes growth of the intestine specifically in mated females, and enhances their reproductive output. The active ovaries of the fly produce the steroid hormone ecdysone, which stimulates the division and expansion of intestinal stem cells in two distinct proliferative phases via the steroid receptors EcR and Usp and their downstream targets Broad, Eip75B and Hr3. Although ecdysone-dependent growth of the female gut augments fecundity, the more active and more numerous intestinal stem cells also increase female susceptibility to age-dependent gut dysplasia and tumorigenesis, thus potentially reducing lifespan. This work highlights the trade-offs in fitness traits that occur when inter-organ signalling alters stem-cell behaviour to optimize organ size.

Steroidal sex hormones including oestrogen, progesterone and testosterone regulate the growth and physiology of reproductive organs during puberty, the oestrus cycle and pregnancy. Consequently, these hormones also promote tumorigenesis in the breast, uterus and prostate. Although sex-specific differences in physiology and disease predisposition extend to nearly all organs[4], the functions of sex-specific steroids in non-sex organs remain relatively poorly explored and controversial. *Drosophila* uses one major steroid hormone, 20-hydroxy-ecdysone (ecdysone, also known as 20HE) and its derivatives[5,6]. Similar to vertebrate steroids, 20HE is synthesized by cytochrome P450 enzymes from cholesterol. The ecdysone receptor comprises a ligand-binding EcR subunit and a DNA-binding Usp subunit—orthologues of human farnesoid X and liver X receptors (FXR and LXR) and retinoid X receptor (RXR), respectively. In juvenile insects, 20HE regulates developmental transitions including moulting, metamorphosis and sexual maturation. In adult *Drosophila*, 20HE is made by the ovaries after mating, resulting in higher levels in females than in males[3,5,7]. It acts in the adult nervous and reproductive systems[3,8] and affects metabolism and lifespan[9,10], but a role in the gut has not been described.

Drosophila intestinal stem cells (ISCs) are more proliferative in females than in males, and females are more prone to age-dependent gut dysplasia and intestinal tumours[2,11,12]. These sex-specific traits could be due to ISC-autonomous and/or systemic factors. Consistent with the former, stress-dependent ISC divisions, which are more frequent in females than in males[2] (Extended Data Fig. 1a, b), are reduced if the ISCs are masculinized by repressing the sex-determination genes *sxl* or *tra*[2] (Fig. 1a, Extended Data Fig. 1b). Mated females support more ISC division than virgin flies (Extended Data Fig. 1a–c), which suggests hormonal influences. Because mated females have higher titres of ecdysteroid than virgins or males[3,5,7], we tested whether 20HE might affect ISC proliferation. Indeed, feeding virgin females 5 mM 20HE strongly induced ISC divisions. This effect was independent of ISC sex identity (Fig. 1a, Extended Data Fig. 1d), and also occurred in mated females and males (Fig. 1b–d, Extended Data Fig. 1a). Using reporters of receptor activity, we confirmed that exogenous 20HE promotes EcR–Usp signalling in midgut ISCs, transient progenitors known as enteroblasts (EBs) and differentiated absorptive enterocytes (ECs) (Extended Data Fig. 1f–j).

Unlike stress caused by detergents, 20HE treatment induced two successive waves of ISC division (Fig. 1d, Extended Data Fig. 1e). Using RNA interference (RNAi) under the control of conditional cell-type-specific *Gal4* drivers, we found that the first wave (at 6 h after 20HE feeding) required EcR only in ISCs (Fig. 1e), but that later divisions (at 16 h) also depended partially on EcR in EBs (Fig. 1e, Extended Data Fig. 2a–f). Neither wave of division required EcR in ECs, enteroendocrine or neural cells (Extended Data Fig. 2g–i). Isoform-specific tests revealed that EcR-A was much more important than EcR-B for the 20HE-induced division of ISCs (Extended Data Fig. 2k–m). 20HE-induced divisions were reversible (Extended Data Fig. 2j), which suggests a lack of toxicity. EcR activity was not induced by enteric infection (Extended Data Fig. 1f–h), and EcR was dispensable for infection-induced gut regeneration (Fig. 1h, Extended Data Fig. 2a, k, l, n–q), which indicates a distinct role for EcR in the gut. Loss of Usp, however, did block infection-induced ISC divisions, which suggests that Usp has EcR-independent functions (Fig. 1h, Extended Data Fig. 2a, n, p).

Next we asked whether ISC activation by 20HE involves the Upd–Jak–Stat or Egfr–ERK signalling pathways, which are known to

---

[1]German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), Heidelberg, Germany. [3]Heidelberg University, Heidelberg, Germany. [4]Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [5]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. ✉e-mail: a.teleman@dkfz-heidelberg.de; bruce.edgar@hci.utah.edu

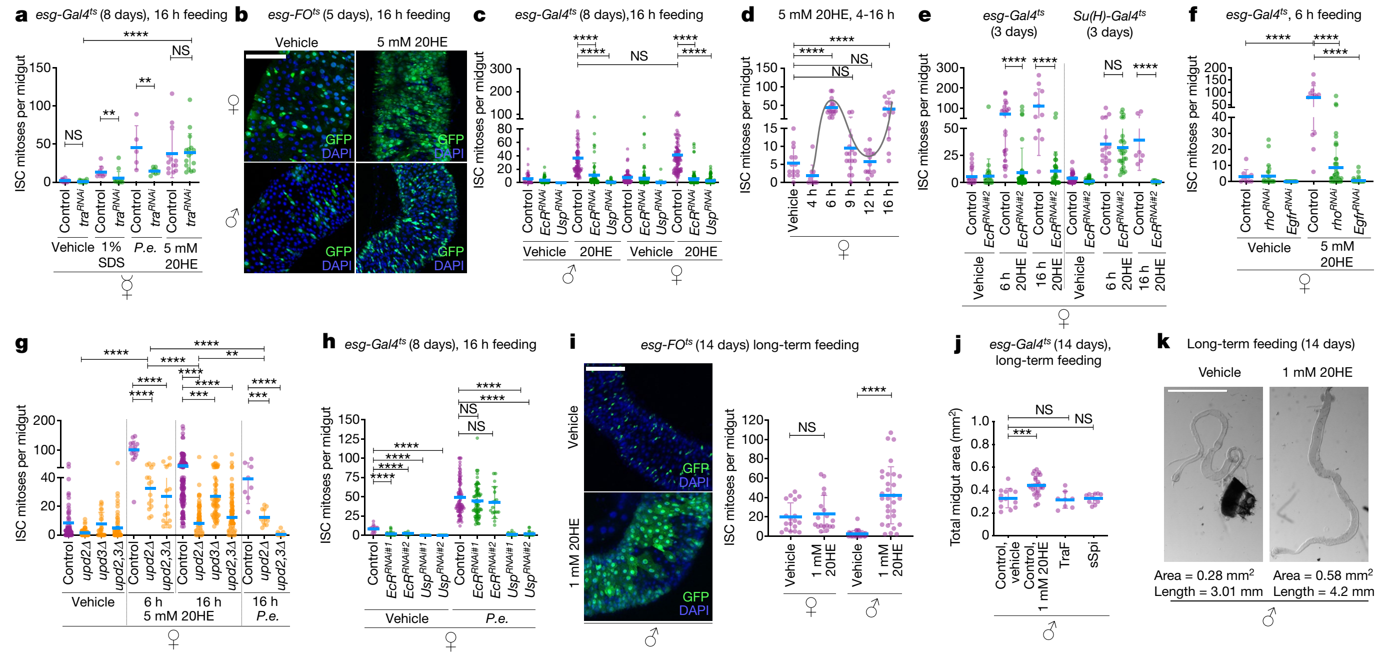
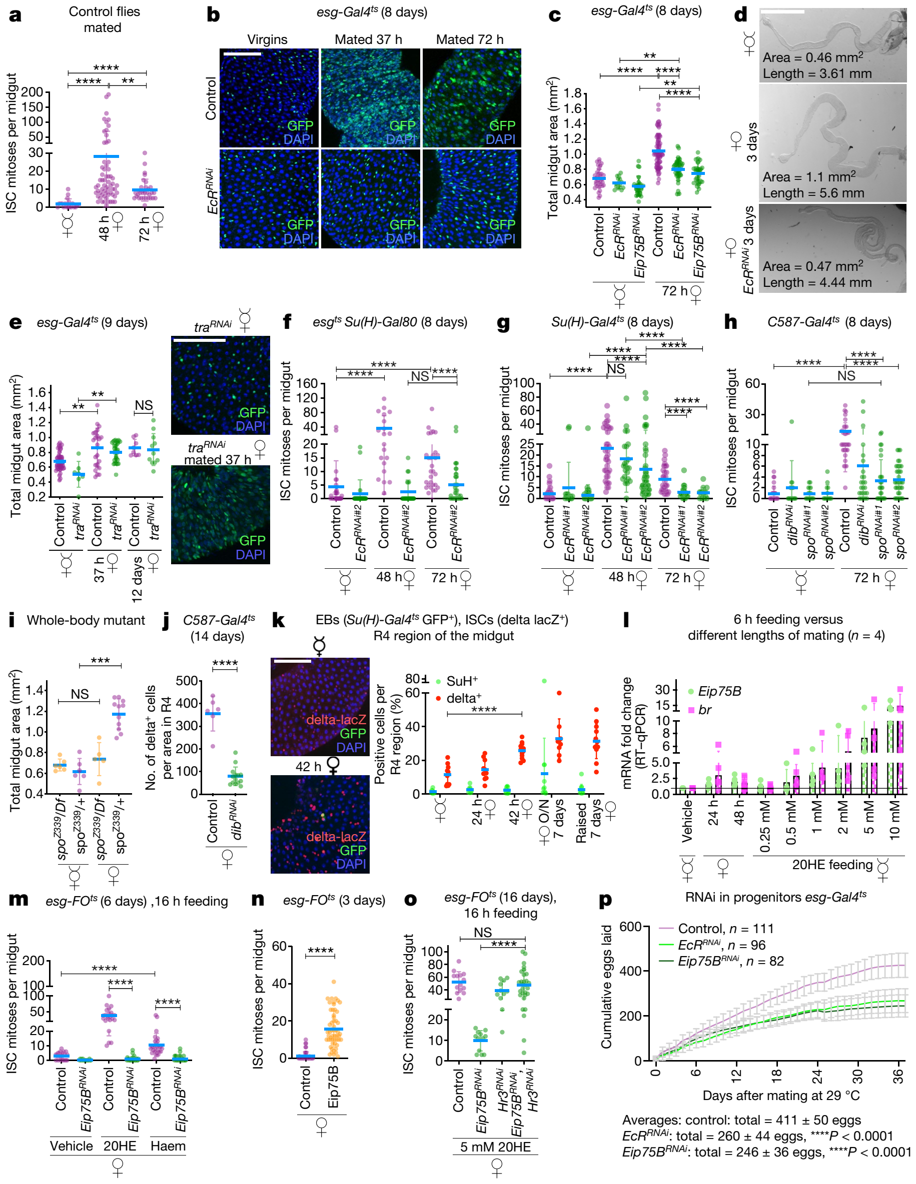


Fig. 1 | Ecdysone induces ISC proliferation and gut growth. a, Midgut mitotic counts of *esg-Gal4*^{ts}>*UAS-tra*^{RNAi} virgins after overnight infection with *Pseudomonas entomophila* (*P.e.*), and feeding with 5 mM 20HE or 1% SDS for 16 h. The *esg-Gal4*^{ts} driver (*esg-Gal4 tub-Gal80*^{ts}) activates UAS target gene expression specifically in ISC and EBs ('progenitor' cells). **b**, ISC lineage-tracing using *esgFO*^{ts}, which drives UAS target gene expression in progenitor cells and their newborn progeny (ECs or EBs) after a temperature shift. **c**, Midgut mitotic counts of *esg-Gal4*^{ts}>*EcR*^{RNAi} and *esg-Gal4*^{ts}>*Usp*^{RNAi} flies fed 5 mM 20HE for 16 h. **d**, Midgut mitotic counts from *w*^{1118} controls fed 5 mM 20HE for different durations. **e**, ISC mitoses in midguts expressing *EcR*^{RNAi} in ISCs and EBs (left) or in EBs (right) after 6 or 16 h of 5 mM 20HE feeding. **f**, Mitotic counts in midguts expressing *rho*^{RNAi} or *Egfr*^{RNAi} in ISCs and EBs 6 h after 20HE feeding. **g**, Midgut mitotic counts of *upd2Δ*, *upd3Δ* and *upd2,3Δ* 5–7-day-old mutant and control flies after 6 and 16 h of 5 mM 20HE feeding or *P.e.* infection. **h**, Mitoses of *EcR*^{RNAi}- or *Usp*^{RNAi}-expressing ISCs and EBs after *P.e.* infection. **i**, Lineage-tracing experiment using *esgFO*^{ts} and mitotic counts in 20HE-fed flies. **j**, Midgut areas from male midguts expressing feminizing TraF or sSpi, or fed 20HE for 14 days. **k**, Male midgut images. Data are mean ± s.d. NS (not significant), $P > 0.05$, $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, Mann–Whitney (**a**, **c**–**i**) or ordinary analysis of variance (ANOVA) (**j**) tests, followed by Bonferroni's multiple comparisons test. Exact *n* values and *P* values are in the Source Data. $n \geq 3$ independent experiments. Scale bars, 100 µm (**b**, **i**) or 1 mm (**k**). DNA counterstained blue with DAPI. In all figures, ♂ denotes males, ☿ denotes virgin females, and ♀ denotes mated females.

activate ISCs after stress[11,13]. Six hours of 20HE feeding induced the Egfr ligands *spi* and *krn* and their activating protease *rho*, but not the *upd2* or *upd3* cytokines or Stat signalling (Extended Data Fig. 3a, b). Exposure to 20HE for 16 h, however, moderately induced *upd2*, *upd3* and Stat activity (Extended Data Fig. 3c–h). The induction of *upd2*, *upd3* and *rho* required EcR in ISCs and EBs (that is, 'progenitors'), although not in ECs (Extended Data Fig. 3c–e). The Egfr effector ERK was also mildly activated by 16 h of 20HE exposure, mostly in progenitors but occasionally in ECs (Extended Data Fig. 3i). ERK activation required *upd2* (Extended Data Fig. 3i), which suggests a signalling relay[13,14]. Notably, the induction of all of these targets (*upd2*, *upd3*, *Socs36E*, *rho*, *spi* and *krn*) by 20HE was suppressed by blocking ISC mitoses with RNAi molecules that target *string* (also known as *stg* or *cdc25*) or Egfr (Extended Data Fig. 3f). This suggests that the observed increases in Jak–Stat and Egfr–ERK signalling are responses to epithelial stress from the early ISC divisions[14]. In further tests, we found that Upd2 from EBs and ECs contributed strongly to ISC divisions 16 h after 20HE feeding, but only weakly to the early divisions at 6 h (Fig. 1g, Extended Data Fig. 3j–l). Egfr and Rho, however, were always required (Fig. 1f, Extended Data Fig. 3j–m). We conclude that ISC divisions are initially activated ISC-autonomously via EcR, and require Egfr and Rho, whereas later divisions depend in part on cytokines produced by EBs and ECs, perhaps in response to stress from the first mitoses. The relationship of EcR to Egfr signalling warrants further investigation.

Because mated females produce more ecdysone than virgins or males[3,5,7], we tested whether 20HE might account for sex-specific differences in the gut. Consistent with this, long-term exposure of males to 20HE phenocopied the female condition, increasing ISC mitoses, stress responsiveness, epithelial turnover and midgut size (Fig. 1i–k, Extended Data Fig. 4a–c). Genetically feminizing the male ISCs did not give these effects (Fig. 1j), which suggests that 20HE acts independently of genetic sex determination. Forced expression of the ISC mitogen[13] sSpi also failed to enlarge male midguts (Fig. 1j), which indicates that 20HE affects more than just the ISC mitotic rate. Long-term 20HE feeding also endowed ISCs in virgin females with proliferative characteristics similar to those seen after mating (Extended Data Fig. 4d). By contrast, RNAi lines that antagonized 20HE signalling in ISCs and EBs decreased gut size in mated females and suppressed mitoses in response to detergent stress (Fig. 2c, d, Extended Data Fig. 4e–g). Thus, sexually dimorphic proliferative traits of ISCs are determined in part by 20HE signalling.

Similar to human oestrogen and progesterone, ecdysone promotes behavioural and metabolic changes that enhance female reproduction[3,7,8]. Dose–response assays showed that 1 mM 20HE fed to virgin females activated EcR targets and ISC mitoses to similar levels to mating (Fig. 2l, Extended Data Fig. 5a). Hence, we tested whether endogenous, mating-induced 20HE activates ISCs. Indeed, mating induced a large, transient increase in ISC division and enduring gut enlargement[15] (Fig. 2a–d, Extended Data Fig. 5b–h, k). This was independent of genetic sexual identity (Fig. 2e, Extended Data Fig. 5i). As with exogenously fed 20HE, these effects initially required EcR only in ISCs, although EcR in EBs contributed later on (Fig. 2f, g, Extended Data Fig. 5e–j). Similar to exogenous 20HE, mating also induced expression of *upd2* and *rho* (Extended Data Fig. 5l), which suggests that these are normal physiological responses.



Averages: control: total = 411 ± 50 eggs
EcR^{RNAi}: total = 260 ± 44 eggs, ****$P < 0.0001$
$Eip75B^{RNAi}$: total = 246 ± 36 eggs, ****$P < 0.0001$

Fig. 2 | Ovary-derived 20HE drives mating-induced midgut growth through Eip75B. a, Midgut mitotic counts from control w^{1118} flies before mating and 48 and 72 h thereafter. **b**, Representative virgin and mated midguts, with their ISCs and EBs marked by GFP, with or without EcR depletion using EcR^{RNAi}. **c**, Midgut size measurements of virgin and mated females, with or without EcR^{RNAi} or $Eip75B^{RNAi}$ expressed in progenitors. **d**, Midgut images with or without EcR-depleted progenitors. **e**, Midgut size measurements and images of virgin and mated females with masculinized tra^{RNAi}-treated ISCs and EBs. **f**, ISC mitoses of virgin female midguts before and 48 or 72 h after mating expressing EcR^{RNAi} specifically in ISCs using the esg^{ts} *Su*(*H*)-*Gal80* Gal4 driver. **g**, ISC mitoses of virgin female midguts and 48 h or 72 h after mating expressing EcR^{RNAi} in EBs using the EB-specific *Su*(*H*)-$Gal4^{ts}$ driver. **h**, Midgut mitoses after depletion of ecdysone synthesis enzymes Dib or Spo by RNAi using the ovary-specific driver $C587$-$Gal4^{ts}$. **i**, Midgut areas of whole-body *spo* mutants (spo^{Z339}/Df) rescued to adulthood by an exogenous 20HE pulse given to embryos, and controls ($spo^{Z339}/+$). **j**, Ovary-derived 20HE promotes gut ISC number. **k**, Images (left) and counts (right) of the percentage of delta⁺ cells or EBs as a fraction of the total cells per region in R4. O/N, overnight. **l**, Quantitative PCR with reverse transcription (qRT–PCR) analysis of *Eip75B* and *br* mRNA in whole, 20HE-fed virgin or mated female midguts. **m**, Mitoses of $Eip75B^{RNAi}$ ISC clones with or without haem or 20HE treatment. **n**, Mitotic counts of ISC clones overexpressing Eip75B. **o**, Epistasis tests assaying interactions between Eip75B and Hr3. **p**, Cumulative eggs laid by mated females with progenitor-specific EcR^{RNAi} or $Eip75B^{RNAi}$, and controls. Data are mean ± s.d. *$P \leq 0.05$, **$P \leq 0.01$, ***$P < 0.001$, ****$P < 0.0001$, ordinary ANOVA test, followed by Bonferroni's multiple comparisons test (**c**, **e**), Mann–Whitney (**a**, **f**–**o**), or general linear models (GLMs) with binomial errors (**p**). Representative images are shown, $n \geq 3$ independent experiments. Scale bars, 100 μm (**b**, **e**, **k**) or 1 mm (**d**). Exact *n* values and *P* values are in the Source Data.

To confirm the source of endogenous ecdysone, we used ovary-specific *Gal4* drivers[3,7] to express RNAi transgenes that target the ecdysone synthesis enzymes Dib or Spo. This suppressed mating-induced ISC divisions and midgut growth, both of which could be restored by exogenous 20HE (Fig. 2h, j, Extended Data Fig. 5n–p). *spo* mutants[16] also failed to resize the midgut after mating (Fig. 2i, Extended Data Fig. 5m), confirming these results. To learn how the gut grows in mated females, we investigated the effects on cell size and number. Depleting EcR in ECs did not reduce EC size (Extended Data Fig. 5q), but mating caused a large 20HE- and EcR-dependent increase in female ISC numbers (Fig. 2k, Extended Data Fig. 5r–t). This expansion of the stem-cell pool could cause an increase in the total number of midgut cells. These results indicate that mating-dependent ISC division, ISC expansion and gut growth

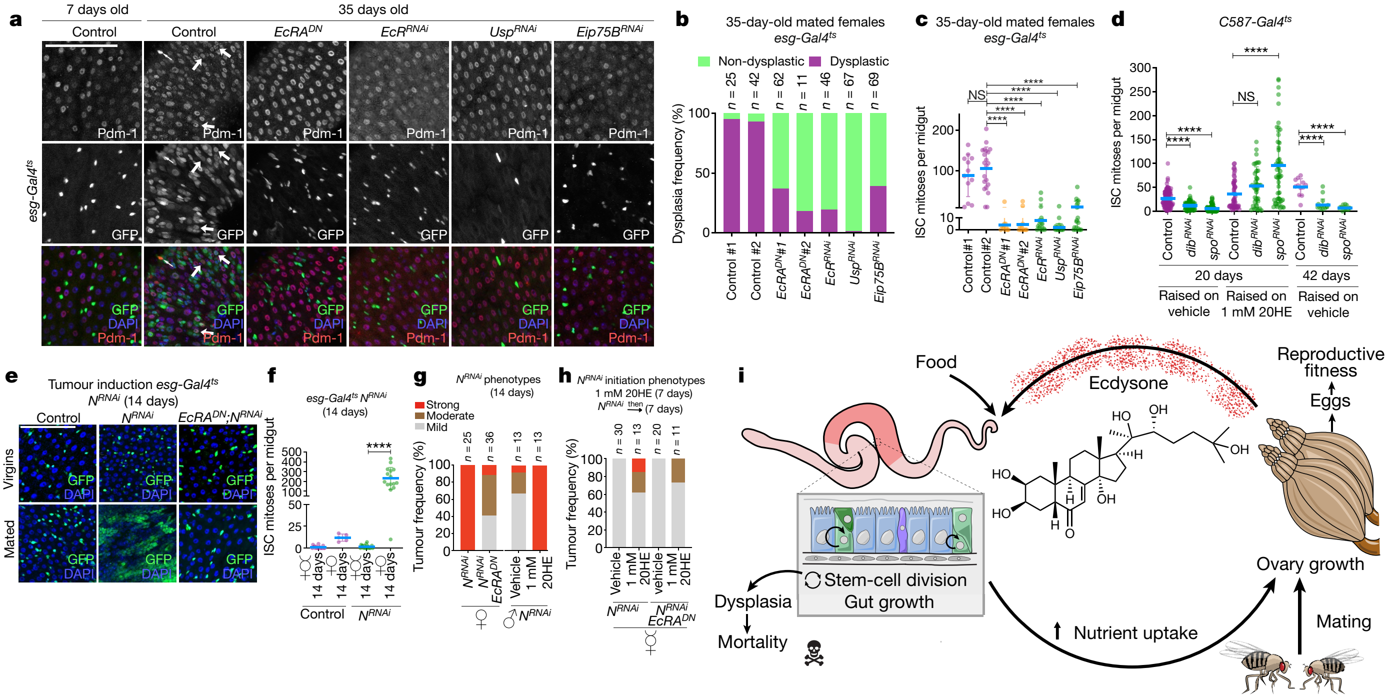


Fig. 3 | Exogenous and mating-dependent 20HE promote intestinal dysplasia and tumorigenesis. a, In aged, control females, mis-differentiated cells accumulate that are positive for both *esg* (GFP, green) and the EC marker Pdm-1 (red; thick white arrows), and can divide (as measured by positive staining for phospho-histone H3 (PH3$^+$); thin white arrows). Intestinal dysplasia is blocked by reduced EcR (using *EcRA^{DN}*), Usp or Eip75B (using *Usp^{RNAi}* or *Eip75B^{RNAi}*), such that progenitors express GFP only. **b**, The percentage of midguts classified as dysplastic (purple) or non-dysplastic (green) in region R4. Guts with at least 10 Pdm1$^+$ GFP$^+$ progenitors were scored as dysplastic. **c**, Mitotic counts of midguts as in **a**. **d**, Midgut mitoses in females expressing ovary-targeted RNAi against the ecdysteroidogenic enzymes Dib or Spo. **e**, *Notch^{RNAi}* (*N^{RNAi}*)-expressing ISC tumours (green cells) are induced in mated females by *esg-Gal4^{ts}*, but not in age-matched virgins. Tumour progression is blocked by *EcRA^{DN}*. **f**, Mitotic counts as in **e**. **g**, Distributions of *esg-Gal4^{ts} UAS-N^{RNAi}* tumour types under conditions as classified in Extended Data Fig. 10d. **h**, Tumour distributions of virgin females fed 1 mM 20HE for 7 days before *N^{RNAi}* induction by *esg-Gal4^{ts}*. Representative images are shown from three independent experiments. Scale bars, 100 μm. **i**, Model summarizing our conclusions. The R4 region is in red. Data are mean ± s.d. ****$P < 0.0001$, Mann–Whitney test. $n \geq 3$ independent experiments. Exact *n* values and *P* values are in the Source Data.

are driven by 20HE signalling from the ovaries to progenitor cells in the gut.

Gut growth after mating is expected to increase the absorption of nutrients by the intestine and the supply of nutrients to other organs. Because egg production is limited by nutrient availability to the ovaries, we tested whether 20HE-dependent gut growth affected female fecundity. When we blocked gut resizing by expressing *EcR* RNAi (*EcR^{RNAi}*) in midgut ISCs, or in both ISCs and EBs, egg production was reduced by approximately 40% (Fig. 2p, Extended Data Fig. 6b–d; see also ref. [7]). This suggests that 20HE-dependent gut remodelling maximizes female reproductive fitness. However, we also noticed that our *Gal4* drivers were active in a small number of escort cells in the germarium of the ovary (Extended Data Fig. 6a, e–l), which raises the possibility that these fecundity defects were due in part to a requirement for EcR in those cells.

A study of *Drosophila* juvenile hormone (JH), a sesquiterpenoid, came to conclusions similar to ours—namely that JH promotes mating-dependent gut growth and fecundity in females[15]. We therefore investigated the relative roles of 20HE and JH. We found that the JH receptors Gce and Met are essential for ISC divisions in response to not only the JH receptor agonist, methoprene[15], but also 20HE and infection (Extended Data Fig. 7a–c). We confirmed the mitogenic effects of methoprene, but these were weaker than those of 20HE (Extended Data Figs. 5a, 7a–g) or mating (Fig. 2a). Furthermore, we discovered that methoprene-stimulated divisions require 20HE (Extended Data Fig. 7g), and that JH or methoprene could suppress ISC divisions driven by 20HE or other stimuli (Extended Data Fig. 7a, d–f). Although these results indicate interaction between 20HE and JH, further work is required to determine their precise physiological relationship (see Supplementary Discussion).

To better understand how ecdysone activates ISCs, we tested two known EcR targets: the transcription factor Broad, and the nuclear receptor Eip75B, a homologue of human PPARγ and REV-ERB[17]. *Eip75B* and *broad* (*br*) mRNA were induced in midguts by 20HE or mating (Fig. 2l, Extended Data Fig. 8a), and progenitor-cell-specific depletion of either factor suppressed 20HE-induced mitoses (Fig. 2m, Extended Data Fig. 8b–e). ISC clonal growth, however, required *Eip75B* but not *br* (Extended Data Figs. 2b, c, 8e, f), highlighting that Eip75B is a more essential effector. Overexpression of Eip75B was sufficient to promote ISC division and gut epithelial turnover (Fig. 2n, Extended Data Fig. 8g), whereas Eip75B loss impaired both ISC mitoses and maintenance (Extended Data Fig. 5s, 8c, f, h, i). Progenitor-specific loss of Eip75B also blocked gut growth after mating (Fig. 2c), and compromised egg production (Fig. 2p, Extended Data Fig. 6b–d), phenocopying the effects of EcR loss. Eip75B binds DNA to repress target genes, and also binds the nuclear receptor Hr3 to inhibit Hr3-mediated transcriptional activation[18]. Consistent with this mechanism, overexpression of Eip75B or 20HE feeding suppressed an Hr3 activity reporter, and Hr3 overexpression suppressed ISC proliferation (Extended Data Fig. 8j–l). Moreover, depletion of Hr3 counteracted losses in ISC proliferation caused by Eip75B depletion (Fig. 2o, Extended Data Fig. 8n). Although these results indicate that Hr3 is a crucial Eip75B effector, Hr3 loss was not sufficient to activate ISC division (Extended Data Fig. 8m, n), which indicates that Eip75B has other targets. Further tests revealed that Eip75B and Hr3 mediate 20HE-independent ISC responses to stress. Enteric infection strongly induced levels of *Eip75B* mRNA (Extended Data Fig. 8a) and suppressed Hr3 activity (Extended Data Fig. 8j). Removing Eip75B or Broad from ISCs by mutation (Extended Data Figs. 2b, c, 8e) or by RNAi (Extended Data Figs. 8b, c, h, i, 9a–d) blocked infection-induced ISC mitoses, as did overexpression of Hr3 (Extended Data Fig. 9e). Eip75B was also required for ISC mitoses in response to the oxidative stress agent paraquat (Extended Data Fig. 8h, i). Furthermore, we obtained evidence consistent with previous work[17] that the action of Eip75B is modulated by haem (a Eip75B ligand) and nitric oxide (Fig. 2m, Extended Data Fig. 9f, g). Functions for haem and nitric oxide in the fly gut are unknown, but potentially interesting. We conclude that Eip75B, Broad and Hr3 integrate several inputs in addition to 20HE to control ISC proliferation (Extended Data Fig. 9h).

As females age, they experience progressive gut dysplasia in which ISCs overproliferate and mis-differentiate, leading to high microbiota loads (dysbiosis), barrier breakdown and decreased lifespan[19,20]. Age-dependent intestinal dysplasia is more pronounced in females than in males[12], and can be identified by increases in mitoses and mis-differentiated cells doubly positive for ISC and EC markers. Suppressing *EcR*, *Usp* or *Eip75B* in midgut progenitors significantly reduced both parameters of dysplasia in aged flies (Fig. 3a–c, Extended Data Fig. 10a). Similarly, suppressing ecdysone synthesis enzymes (Dib, Spo) in the ovaries, or ubiquitously, also curtailed age-dependent gut dysplasia (Fig. 3d, Extended Data Fig. 10b, c). This effect could be reversed by 20HE supplementation. These results indicate that age-dependent gut dysplasia is potentiated by ovary-derived ecdysone, explaining the sex bias of this condition.

Female *Drosophila* are known to be more susceptible than males to genetically induced ISC-derived tumours. We found that ISC/EB-specific RNAi targeting *Notch*, a receptor required for EC differentiation, drove tumour induction in 100% of mated females but was far less tumorigenic in virgin females or males (Fig. 3e–g, Extended Data Fig. 10d, e). Three results indicate that this tumour predisposition is modulated by 20HE. First, in contrast to mated females, virgins were extremely resistant to *Notch*^{RNAi}-mediated tumorigenesis (Fig. 3e–h). Second, targeting 20HE signalling in ISCs with a dominant-negative EcR-A (*EcRA*^{DN}) inhibited tumour growth in mated females (Fig. 3e, g, Extended Data Fig. 10d). Third, supplementing males or virgin females with 20HE increased tumour initiation and growth (Fig. 3g, h, Extended Data Fig. 10f). We surmise that the use of mating-dependent, ovary-derived 20HE to stimulate gut resizing comes at a cost: it predisposes females to gut dysplasia and tumorigenesis (Fig. 3i, Extended Data Fig. 9i).

Gut dysplasia, tumorigenesis and egg production can all shorten lifespan[10,20,21], which suggests that the effects of ecdysone on the gut might adversely affect longevity. In fact, earlier reports showed that *EcR* mutants live longer[9], and proposed that reproduction can shorten lifespan by damaging the soma[22]. Our own lifespan assays, although subject to the same caveats as previous work[20] (Supplementary Discussion), support this view: suppression of EcR in midgut progenitors extended lifespan in females but not males (Extended Data Fig. 10g–i). In evolutionary terms, the disadvantage of a slightly shorter lifespan due to sex-specific hormonal signalling is probably insignificant relative to the reproductive fitness advantage conferred by increased egg production. This may be especially true in the wild, where gut dysplasia-dependent mortality is probably counteracted by nutrient deprivation[12].

Similarities in the reproductive biology of *Drosophila*[3,8] and mammals[23] suggest that these inter-organ relationships have relevance to human biology. The mitogenic effects of insect ecdysone parallel those of oestrogen and testosterone as drivers of breast, uterine and prostate growth and tumorigenesis. Yet how these steroids affect the human intestine remains poorly explored. Adaptive growth of the intestine is well documented in pregnant and lactating mammals[24], and might depend on oestrogen and/or progesterone. Laboratory tests with rodents and human cells, as well as some studies with human participants, have linked oestrogen, testosterone and their receptors to gastrointestinal cancers[25], but epidemiological studies provide conflicting evidence regarding this association[26,27] (Supplementary Discussion). The contributions of sex steroids to intestinal physiology deserve more detailed study.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2462-y.

1. Capel, B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet*. **18**, 675–689 (2017).
2. Hudry, B., Khadayate, S. & Miguel-Aliaga, I. The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature* **530**, 344–348 (2016).
3. Sieber, M. H. & Spradling, A. C. Steroid signaling establishes a female metabolic state and regulates SREBP to control oocyte lipid accumulation. *Curr. Biol*. **25**, 993–1004 (2015).
4. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat. Rev. Genet*. **9**, 911–922 (2008).
5. Schwedes, C. C. & Carney, G. E. Ecdysone signaling in adult *Drosophila melanogaster*. *J. Insect Physiol*. **58**, 293–302 (2012).
6. Lavrynenko, O. et al. The ecdysteroidome of *Drosophila*: influence of diet and development. *Development* **142**, 3758–3768 (2015).
7. Ameku, T. & Niwa, R. Mating-induced increase in germline stem cells via the neuroendocrine system in female *Drosophila*. *PLoS Genet*. **12**, e1006123 (2016).
8. Meiselman, M. et al. Endocrine network essential for reproductive success in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **114**, E3849–E3858 (2017).
9. Simon, A. F., Shih, C., Mack, A. & Benzer, S. Steroid control of longevity in *Drosophila melanogaster*. *Science* **299**, 1407–1410 (2003).
10. Tricoire, H. et al. The steroid hormone receptor EcR finely modulates *Drosophila* lifespan during adulthood in a sex-specific manner. *Mech. Ageing Dev*. **130**, 547–552 (2009).
11. Jiang, H. et al. Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* **137**, 1343–1355 (2009).
12. Regan, J. C. et al. Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *eLife* **5**, e10956 (2016).
13. Jiang, H., Grenley, M. O., Bravo, M. J., Blumhagen, R. Z. & Edgar, B. A. EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in *Drosophila*. *Cell Stem Cell* **8**, 84–95 (2011).
14. Zhang, P. et al. An SH3PX1-dependent endocytosis-autophagy network restrains intestinal stem cell proliferation by counteracting EGFR-ERK signaling. *Dev. Cell* **49**, 574–589 (2019).
15. Reiff, T. et al. Endocrine remodelling of the adult intestine sustains reproduction in *Drosophila*. *eLife* **4**, e06930 (2015).
16. Ono, H. et al. Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Dev. Biol*. **298**, 555–570 (2006).
17. Marvin, K. A. et al. Nuclear receptors *Homo sapiens* Rev-erbβ and *Drosophila melanogaster* E75 are thiolate-ligated heme proteins which undergo redox-mediated ligand switching and bind CO and NO. *Biochemistry* **48**, 7056–7071 (2009).
18. White, K. P., Hurban, P., Watanabe, T. & Hogness, D. S. Coordination of *Drosophila* metamorphosis by two ecdysone-induced nuclear receptors. *Science* **276**, 114–117 (1997).
19. Biteau, B., Hochmuth, C. E. & Jasper, H. JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell* **3**, 442–455 (2008).
20. Biteau, B. et al. Lifespan extension by preserving proliferative homeostasis in *Drosophila*. *PLoS Genet*. **6**, e1001159 (2010).
21. Patel, P. H., Dutta, D. & Edgar, B. A. Niche appropriation by *Drosophila* intestinal stem cell tumours. *Nat. Cell Biol*. **17**, 1182–1192 (2015).
22. O'Brien, D. M., Min, K.-J., Larsen, T. & Tatar, M. Use of stable isotopes to examine how dietary restriction extends *Drosophila* lifespan. *Curr. Biol*. **18**, R155–R156 (2008).
23. Speakman, J. R. The physiological costs of reproduction in small mammals. *Phil. Trans. R. Soc. Lond. B* **363**, 375–398 (2008).
24. Hammond, K. A. Adaptation of the maternal intestine during lactation. *J. Mammary Gland Biol. Neoplasia* **2**, 243–252 (1997).
25. Amos-Landgraf, J. M. et al. Sex disparity in colonic adenomagenesis involves promotion by male hormones, not protection by female hormones. *Proc. Natl Acad. Sci. USA* **111**, 16514–16519 (2014).
26. Manson, J. E. et al. Menopausal hormone therapy and long-term all-cause and cause-specific mortality: the women's health initiative randomized trials. *J. Am. Med. Assoc*. **318**, 927–938 (2017).
27. Gunter, M. J. et al. Insulin, insulin-like growth factor-I, endogenous estradiol, and risk of colorectal cancer in postmenopausal women. *Cancer Res*. **68**, 329–337 (2008).

Methods

Drosophila stocks and cultures

Drosophila melanogaster were raised on standard media and maintained in incubators with controlled temperature and humidity on a 12 h light/dark cycle. Flies were transferred to fresh vials every 2 days. Male and female *Drosophila* were raised mated for all experiments, unless otherwise indicated. To generate controls, *w*^{1118} (VDRC 60000) flies were typically outcrossed to the appropriate *Gal4* driver line. To generate controls for experiments using VDRC 'KK' RNAi lines, the stock *y w[1118]; P{attP,y[+],w[3`]* (VDRC 60100) was outcrossed to the appropriate *Gal4* driver line. Full genotypes of all stocks used, and for each figure panel, are listed in Supplementary Tables 1 and 2.

Drosophila husbandry

For transgene expression using the *Gal4/Gal80*^{ts} system, experimental crosses were maintained at 18 °C (permissive temperature for *GAL80*^{ts}) in standard medium. Animals of the desired sex and genotype were collected within 48 h of eclosion and aged for an average of 5 days before shifting to 29 °C (restrictive temperature for *GAL80*^{ts}) to induce UAS transgene expression. Adult midguts were dissected after different periods of time as indicated in each figure. The esg-Flip-Out system (*esgFO*^{ts})[28] and the MARCM system[29] were used to generate ISC-derived clones. Flies were aged for 3–6 days after eclosion before clonal induction by temperature shift to 29 °C for *esgFO* clones or heat-shock for MARCM clones. Further details on transgene expression times are indicated in the figure legends. MARCM 80B flies were heat-shocked for 45–60 min in a 37 °C water bath, and then aged for 12 days at 29 °C before overnight treatment with vehicle, 5 mM 20HE or *Psuedomonas entomophila*.

Mating experiments

At least 10–15 virgin females for each genotype were collected at 18 °C as they emerged. They were aged for approximately 5 days and then shifted to 29 °C until the time points indicated in each figure. At the start of mating, females were transferred to fresh vials and allowed to mate with equal numbers of adult 3–7 days old wild type *w*^{1118} males, devoid of any transgenes, at 25 °C, for optimal fecundity. Time when males were introduced to females in the same vial is denoted as *t*0. If indicated as mated once, then after 18–20 h, the males were removed and the females were flipped into fresh vials every 48 h until the indicated time in the respective figures. Otherwise, males were left together with the females for the following time points: 24 h, 37–40 h, 46–48 h or 72–74 h.

GAL4-LBD 'ligand sensor' system

Adult flies with bipartite detection system consisting of the LBD of the *Drosophila* nuclear receptor fused to the DNA-binding domain of yeast GAL4, along with a GAL4 UAS-controlled GFP reporter gene were used as previously described[30,31]. Flies were raised and maintained at 25 °C. For visualization of ligand sensor patterns, 5–7-day-old mated females were starved for 2–4 h, heat-treated for 30 min in a 37 °C water bath only once for EcR, Usp and Hr3 reporters, and allowed to recover at room temperature for 15 min. Then, flies were transferred to vials containing a fresh feeding vial (see 'Feeding experiments') and kept at 25 °C for 16–18 h until dissection.

In vivo 10XSTAT92E-GFP reporter system

Adult mated female flies of the genetic background 10XSTAT92E-GFP that have 10 Stat92E-binding sites driving GFP expression were aged for 5–7 days and treated for 6 h with 5 mM 20HE and for 16–18 h with 5 mM 20HE or *P. entomophila* infection.

In vivo upd3-lacZ reporter

Adult mated female flies of the genetic background Upd3.1 LacZ/TM6B were aged for 5–7 days and treated for 16–18 h with 5 mM 20HE or *P. entomophila* infection.

Overnight feeding experiments

For all experiments except 20HE or SDS feeding (as indicated in the figures), flies were fed for 16–20 h, then dissected to remove the intestines, which were analysed using immunofluorescence and confocal imaging. For timed 20HE feeding, flies were collected as early as 4 h and as late as 22 h after continuous 20HE exposure. We observed a window of strong mitotic response at 6 h and again at 16–18 h that persisted to 22 h after exposing the flies to the 20HE feeding solution.

For 20HE removal experiments, flies were fed overnight for 16–18 h with 5 mM 20HE, and then transferred to a fresh vial for another overnight treatment after which the midguts were dissected and stained.

20HE feeding: 10–15 adult male, mated female or virgin female flies were used for the ecdysone feeding experiments, as indicated. 20HE was dissolved in 100% ethanol, water was added to make a 25 mM stock solution in 10% ethanol, and stocks were stored at −20 °C. A final concentration of 0.25–10 mM ecdysone or 2% ethanol (as control) was used for the feeding experiments as indicated. Then, 200 µl of 5% sucrose solution, 5 mg ml^{-1} dry yeast and 5 mM 20HE (Sigma-Aldrich H5142) mix was deposited on top of a standard food vial to which flies were transferred. If the experiment required *P. entomophila* infection, then 400 µl of the same yeast/sucrose mix (described above) was deposited on filter-paper discs (Whatmann) to which flies were being transferred. The sucrose yeast mix with 2% ethanol was used as vehicle treatment.

Detergent treatment: flies were left to feed on yeast sucrose solution (described above) with 0.1% or 1% SDS for 18–20 h or at the times indicated.

Enteric *P. entomophila* infection: a 25 ml pre-culture was started the first day by inoculating *P. entomophila* bacteria from glycerol stocks (stored at −80 °C) in Rifampicin-supplemented Luria Broth (LB; final antibiotic concentration: 100 µg ml^{-1}). The pre-culture was grown overnight at 29 °C, shaking at 130 rpm. The next day, the pre-culture was diluted in 175 ml Rifampicin-supplemented LB and the culture was again grown overnight at 29 °C, shaking at 130 rpm. After the growth of the bacterial culture reached optical density of approximately 0.5, the culture was spun down at 2,500*g* for 25 min at 4 °C and the pellet was re-suspended in 3 ml of 5% sucrose plus 150 µl yeast. Before infection, flies were starved for 2 h (optional step), and then placed in vials with 500 µl of this *P. entomophila* solution or 5% sucrose with yeast as the control vehicle.

Other treatments in Fig. 3 and Extended Data Figs. 8, 9 include feeding with 2.5 mM paraquat, Nω-nitro-L-arginine methyl ester hydrochloride (Sigma-Aldrich, N5751) (200 mM L-NAME stock solution in distilled water; final 10 mM concentration was used), (±)-*S*-nitroso-*N*-acetylpenicillamine (Sigma-Aldrich, N3398) (500 mM SNAP stock solution in 10% ethanol and 10 mM SNAP final solution was used), hemin (Frontier Scientific, H651-9) (2 mM stock solution dissolved in 0.1 M NaOH, pH adjusted to 7 with sodium phosphate buffer and 0.5 mM final solution was immediately used) and their corresponding vehicle. Treatments were diluted in 400 lL total volume of 5% sucrose and 5 mg ml^{-1} yeast then added vials containing a fresh feeding paper.

Long-term ecdysone feeding

At least 10–15 adult male and/or female flies were transferred to standard fresh food vials (2.5 cm diameter) containing circa 3 ml of food. To prepare ecdysone treated food, the food in the vial was scraped on the surface and 200 µl 1 mM 20HE, 22 mg ml^{-1} yeast in 5% sucrose solution was added. After 15 min, this solution diffused into the food. Flies were added to these vials and flipped into fresh 20HE containing vials every 48 h for 14 days unless otherwise indicated. As vehicle, vials with fly media containing 200 µl 0.43% ethanol in sucrose/yeast solution were used. Flies were dissected to remove the intestines, which were analysed using immunofluorescence and confocal imaging. For the flies raised on low nutrient food, flies were fed with 1 mM 20HE, 5 mg ml^{-1} yeast in 5% sucrose solution that was deposited on filter-paper discs (Whatmann) and exchanged every 24–30 h. For *P. entomophila* infection after long-term 20HE feeding in Extended Data Fig. 4c, we discontinued feeding the flies on ecdysone-containing food for one day before the flies were fed with the *P. entomophila* bacterial solution.

Fecundity assays

Fig. 2p, Extended Data Fig. 6b: 10–15 virgin females for each genotype/replicate were collected at 18 °C as they eclosed, and pooled in one vial. For each genotype, 3–4 replicates were performed for every experiment. Virgins were aged one day and then shifted to 29 °C to activate Gal4. Females were then transferred to fresh cages and allowed to mate with equal numbers of w^{1118} males. Females were housed in groups of 7–10 with equal number of males for this experiment. Standard *Drosophila* media was poured in 5 cm plates and stored at 4 °C. Flies in each egg collection cage were flipped onto fresh food plates every 24–48 h for the indicated number of days, and the number of eggs/replicate were scored and averaged over the number of flies in each cage. Three to four independent experiments were performed, all results were pooled, and are shown in Fig. 2p and Extended Data Fig. 6b. Raw egg counts, processed cumulative sums, averages, and *P* values for each experiment are in the Source Data.

Extended Data Figure 6d: virgins were aged for 8 days and shifted to 29 °C to activate Gal4 first before mating to equal number of males. Females were housed in groups of 7–10 with equal number of males for this experiment. Flies in each egg collection cage were flipped onto fresh food plates every 24–48 h for the indicated number of days, the number of eggs/replicate were scored and averaged over the number of flies in each cage. Cages with dead flies were excluded from the analysis. Raw egg counts, processed sums and *P* values are available in the online source data.

Extended Data Figure 6c, d: virgins were aged for one day and then shifted to 29 °C to activate Gal4. Females were then transferred to fresh vials and allowed to mate with equal numbers of w^{1118} males. All subjects were housed overnight in the same vial to ensure mating success and numbers of eggs were counted and averaged for the number of females/vial. Next day, every female and male pair was separated and individual females or vials were followed up for 14 days. Vials were exchanged every 24–48 h in this experiment and the total number of eggs laid every 2 days was counted for every female fly. Vials with dead flies were excluded from the analysis. Raw egg counts, processed cumulative sums/averages and *P* values are in the online Source Data.

A 2- or 3-day sum was calculated from the average number of eggs or flies laid every day, and then an average sum of eggs laid per fly per 3 days across the replicates was plotted with error bars ± confidence intervals. Alternatively, the average or individual cumulative numbers of eggs were summed up and mean values were plotted ± standard deviation. To test statistical significance for each day, two-sample unequal variance *t*-test were performed, with a two-tailed distribution assuming unequal variance for test genotype relative to control at every time point. Individual *P* values are in the online Source Data. Alternatively, for Fig. 2p, general linear models with binomial errors were used to examine the effect of the genotype on the average cumulative number of eggs.

spo mutant rescue experiment

Males of either deficiency backgrounds BM#7584 or BM#24411 were crossed to heterozygous *spo* mutant virgins and allowed to lay eggs on apple plates for several days before the experiment. Two deficiency genotypes were used to increase the likelihood to getting rescued homozygous spo mutant flies. On the day of the experiment, the parents were left to lay eggs for 4 h then, were removed. The eggs were allowed to age 4–6 h at 25 °C then, were all pooled in a sieve and de-chorionated by bleach. After washing in PBS-T, the de-chorionated embryos were incubated in PBS-T supplemented with 100 µM 20E for 3 h. The embryos were covered with Halocarbon 27 oil and incubated at 18 °C overnight. Over the next 2 days, homozygous *spo* embryos were selected under a fluorescent stereoscope by the lack of GFP expression in the hatched larvae. The phenotypically correct larvae were collected in fresh food vials at the density of 40–60 larvae per vial and allowed to develop at 25 °C until eclosion and selection of virgin or mated homozygous spo mutant flies.

Lifespan assays

Males and females of the genotype 5961GS EcR A DN were allowed to mate for 48 h and were then isolated in groups of 25 flies of the same sex per vial. For RU486 food supplementation, 100 µl of a 5 mg ml^{-1} solution of RU486 or vehicle (ethanol 80%) was deposited on top of a food vial and dried for at least 4–6 h, resulting in a 0.2 mg ml^{-1} concentration of RU486 in the food accessible to flies. Flies were flipped every 48 h into a fresh vial. Dead flies were visually identified (flies not moving, not responding to mechanical stimulation and lying on their side or back were deemed dead), and the number of dead flies was recorded. Oasis software was used for data analysis[32]. A log-rank non-parametric test was performed by the software and the *P* values were derived from pairwise comparison with Bonferroni correction as displayed in Extended Data Fig. 10g, h.

Immunohistochemistry and microscopy

Drosophila adult midguts were dissected in 1× PBS and fixed with 4% paraformaldehyde for 30 min at room temperature. For all immunostainings except anti-dpErk, samples were washed with 0.015% Triton-X in PBS three times at room temperature, then permeabilized with 0.15% Triton-X in PBS for 15 min at room temperature with shaking. Then, samples were re-washed and blocked in PBS with 2.5% BSA, 10% normal goat serum and 0.1% Tween-20 (blocking solution) for at least 1 h at room temperature. Midguts were incubated with primary antibody at 4 °C overnight at the following dilutions: chicken anti-GFP (Life Technologies/Molecular Probes, 1:500); rabbit anti-phospho-histone 3 (Merck Millipore 1:1,000); mouse anti-phospho-histone 3 (Cell Signaling, 1:1,000); guinea pig anti-GFP (Teleman Lab, 1:1,000); chicken anti-β-galactosidase (Abcam, 1:1,000).

For the dpERK detection, samples were fixed in 4% paraformaldehyde, dehydrated for 5 min in 50%, 75%, 87.5% and 100% methanol, and rehydrated for 5 min in 50%, 25% and 12.5% methanol in PBST (0.1% Triton X-100 in 1× PBS). After washing in 1× PBST, midguts were blocked in PBS with 2.5% BSA, 10% normal goat serum and 0.1% Tween-20 (blocking solution) for at least 1 h at room tmperature then incubated with rabbit phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) 9101 (Cell Signaling, 1:400) at 4 °C overnight.

After washing, all samples were incubated with secondary antibodies (Alexa 488, 568 or 633, Invitrogen) for more than 2 h at room temperature at a dilution of 1:1,000. All antibody incubations were performed in blocking solution. DNA was stained with 0.5 µg ml^{-1} DAPI (Sigma).

For the plasma membrane cell stain: Freshly dissected midguts were stained with CellMask deep red plasma membrane stain, Thermofisher in 1× PBS at a concentration of 1:1,000 then fixed in 4% formaldehyde and stained with 1x PBS/DAPI according to the manufacturer's instructions.

Ovary staining: one-day-old mated females have been place on active yeast paste for 4–5 days at 29 °C. Ovaries were dissected in PBS, transferred in PBS containing 8% paraformaldehyde and fixed for 10 min at room temperature with mixing. After washes in PBS with 0.15% Triton, ovaries were blocked for 1 h in 0.15% PBST containing 2.5% BSA. The following primary antibodies were incubated at 4 °C overnight in blocking buffer: chicken anti-GFP 1:500, mouse anti-coracle 1:500 (DSHB, C566-9). Ovaries were then washed five times for 5 min in 0.15% PBST and incubated for 1 h 30 min with the following secondary antibodies and dyes in blocking buffer at room temperature: goat anti-chicken488 1:1000, goat anti-mouse568 1:1,000, Hoechst 1:1,000, phalloidine633 1:10,000. After two washes for 10 min in 0.15% PBST, ovaries were mounted on slides with Vectashield. Images have been acquired using a Leica Sp8 confocal microscope and the figures made using Fiji with the ScientiFig plugin.

Imaging: midguts were mounted on glass slides in VectaShield (Linaris). All midgut images were acquired on a Leica TCS SP5II inverted confocal microscope, equipped with HCX Plan APO 20×/1.30 glycerol-immersion (for quantifications) or 40×/1.30 oil-immersion objectives (for representative images/quantifications), using Leica Application Suite (LAS) AF software and processed with Fiji/ImageJ software[33]. Representative images are shown. GFP, in green (native GFP for all genotypes except for the reporter midguts and $Su(H)^+$ cells marked with $Su(H)^{ts}$ driver that were also stained with GFP for better visualization of the signal); DNA: DAPI, in blue. To display images in the figure panels, a *Z*-stack of defined steps for control and test genotypes in a single field was acquired in the R4 region (a region which is bounded by the apex of the midgut tube's most distal 180° turn) as previously described[34]. Images represent maximal intensity projections of the acquired *Z*-stacks. Scale bars are 100 µm in all images, unless otherwise indicated.

Quantifications and statistics

ISC proliferation: mitotic indices were determined by manually counting all PH3-positive cells in entire midguts using Leica DM5000B or Zeiss Axiophot fluorescence microscopes through a 40× objective. Statistical analysis of all the mitotic counts was performed using two-tailed Mann–Whitney test. All dot plot graphs indicating mitoses are showing mean ± s.d. Exact *P* values are provided in the Supplementary Information. Data were from at least three independent experiments.

Quantification of the GFP^+/$delta^+$ cells: *Z*-stacks of both epithelial sides in R4a/b region were imaged at steps of 5.0 µm at 40× then the total number of GFP^+ or $delta^+$ cells were analysed after limiting the particle size to 10–250 µm, circularity 0.00–1.00 and excluding holes after maximal *Z*-projects have been applied.

Quantification of the $delta^+$ and $Su(H)^+$ cells: *Z*-stacks of both epithelial sides in the R4a/b region were imaged by confocal Zeiss LSM 780 Spinning Disc. The total number of $DAPI^+$, $Su(H)^+$ and $delta^+$ cells were automatically segmented and counted using a custom Image J/FIJI macro (Supplementary Data 6). $Su(H)^+$ and $delta^+$ cells were manually recounted and verified and the numbers of each cell type were recorded to derive the percentage cell type to total cell number/stack.

Quantification of cell size: midguts were mounted as previously described and *Z*-stacks of both epithelial sides in the R4a/b were imaged at steps 5.0 µm at 40× then a custom Image J/FIJI macro (Supplementary Data 1) was created to segment the cytoplasm in reference to DAPI nuclear stain and internuclear distances. Area of the cells in micrometre-squared were outputted to Microsoft Excel and a mixed effects two-way ANOVA statistical model was computed to calculate the significance between the different conditions.

Quantification of clonal size: *Z*-stacks of both epithelial sides in the R4a/b were imaged at steps 5.0 µm at 40× then a custom Image J/FIJI macro (Supplementary Data 2) was used to semi-automatically segment and determine the location and size of the GFP^+ clones then the sizes in micrometre-squared were outputted to Microsoft Excel and a mixed effects two-way ANOVA statistical model was computed to calculate the significance between the different conditions.

Quantification of the GFP^+ areas: for analysis of the mating effects, *Z*-stacks of both epithelial sides in R4a/b region were imaged at steps 5.0 µm either at 40× or at 20×. For analysis, the quantification of the area occupied by GFP^+ cells was performed automatically using a custom ImageJ/FIJI macro (Supplementary Data 3). The macro created maximum *Z*-projection of image stacks, median and Gaussian filtering, automatic thresholding and measurement of GFP^+ and gut occupying area. The measurements were exported to Microsoft Excel and the GFP^+/gut area ratio was derived from these values for at least 10 midguts for most experiments.

Quantification of the GFP^+ area/$DAPI^+$ cells: for analysis of the tumour effects (Extended Data Fig. 10e), a fixed median filter was created for each stack, a fixed Gaussian blur value was applied; then the midgut was thresholded for $DAPI^+$ cells and GFP^+ cells; then areas for both were calculated and a ratio was derived. An Image J/FIJI macro was used (Supplementary Data 3).

Data are displayed in scatter plots with the mean ± s.d. for each series of experiments. Data shown are representative of at least two or three independent repeated experiments with similar results. Statistical significance was calculated either by two-tailed Mann–Whitney test without a multiple comparison test. Results were considered to be significantly different at $P < 0.05$. All calculations were performed using the Prism 7.0 software (GraphPad Software).

Gut measurements: after immunofluorescence staining and before mounting, midguts were put on a glass slide and imaged using a Leica M205 FA Stereo Microscope or Stereo Discovery.V8, unmounted guts were imaged at a defined magnification and these images were exported to Fiji for further analysis. Custom Image J/FIJI macros (Supplementary Data 4, 5) were used to threshold each image then measure the area of each midgut. With the distance mapping technique, the midgut length was derived. For the width measurements, a line was drawn. Before quantifying any midgut dimensions, the genotype of each sample was concealed. Samples were randomly analysed then the genotype was revealed only after completing analysis. For statistical analyses of gut sizes, normality test was performed with Shapiro–Wilk normality test and the gut sizes showed normal Gaussian distribution. Thus, statistical significance of gut size measurements was calculated by ordinary ANOVA test, followed by Bonferroni's multiple comparisons test. Data are displayed in scatter plots with the mean ± s.d. Data were plotted from at least three independent repeated experiments with similar results.

All Image J/FIJI macros are available as supplementary online source material (Supplementary Data 1-6), or upon request from the authors.

Sample sizes, randomization and blinding

No statistical methods were used to predetermine sample sizes, but typically between 5 and 20 flies were used per replicate per genotype in each experiment. Exact *n* values for each experiment are in the online Source Data. When selecting animals for an experiment, the parental genotype was not concealed because it was required to select pertinent progeny. Animals were first selected by genotype and then randomly chosen for experimental analysis. For measurements of mitoses/gut, gut sizes and tumour frequencies, the genotype of each sample was concealed during analysis. Samples were then randomly scored and genotypes were revealed only after completing the analysis.

RT–qPCR

Approximately 10–12 female intestines per genotype were dissected and RNA isolated using the RNAeasy kit (QIAGEN). Then, 750 ng of total RNA was used for cDNA synthesis reactions using the QuantiTect reverse transcription kit (QIAGEN). RT–qPCR was performed on a Light Cycler 480 II (Roche) using SYBR Green I (Roche). Experiments were performed in at least biological triplicates. Relative fold differences in expression level of target genes were calculated as ratios to the mean of the reference genes *rp49* and tubulin using the $\Delta\Delta C_t$ method. A series of tenfold dilutions of an external standard was used in each run to produce a standard curve. Primer sequences are listed in Supplementary Table 3.

$\Delta\Delta C_t$ method: $\Delta\Delta C_t$ (or $\log_2$-transformed fold change) is the difference in threshold cycles for the test and control sample normalized to the threshold cycles for the reference gene.

$\Delta\Delta C_t = \Delta C_t\,(\text{test}) - \Delta C_t\,(\text{control})$

$\Delta C_t\,(\text{test})$ or $\Delta C_t\,(\text{control}) = C_t$ target gene $- C_t$ reference gene

All data are presented as mean $\log_2$-transformed fold change with s.d.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source data are provided with this paper.

Code availability

Code for all FIJI macros used in this study is available for download via the Supplementary Information. These macros are available as Supplementary Data 1–6.

28. Jiang, H. & Edgar, B. A. EGFR signaling regulates the proliferation of *Drosophila* adult midgut progenitors. *Development* **136**, 483–493 (2009).
29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
30. Palanker, L. et al. Dynamic regulation of *Drosophila* nuclear receptor activity in vivo. *Development* **133**, 3549–3562 (2006).
31. Kozlova, T. & Thummel, C. S. Spatial patterns of ecdysteroid receptor activation during the onset of *Drosophila* metamorphosis. *Development* **129**, 1739–1750 (2002).
32. Han, S. K. et al. OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research. *Oncotarget* **7**, 56147–56152 (2016).
33. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
34. Liang, J., Balachandra, S., Ngo, S. & O'Brien, L. E. Feedback regulation of steady-state epithelial turnover and organ size. *Nature* **548**, 588–591 (2017).
35. Zheng, W. et al. Dehydration triggers ecdysone-mediated recognition-protein priming and elevated anti-bacterial immune responses in *Drosophila* Malpighian tubule renal cells. *BMC Biol*. **16**, 60 (2018).
36. Su, Y.-H. et al. Diet regulates membrane extension and survival of niche escort cells for germline homeostasis via insulin signaling. *Development* **145**, dev159186 (2018).
37. Cáceres, L. et al. Nitric oxide coordinates metabolism, growth, and development via the nuclear receptor E75. *Genes Dev*. **25**, 1476–1485 (2011).

Acknowledgements We thank J. Zhou, I. Miguel-Aliaga, C. Thummel, P. Patel and L. O'Brien for stocks and discussions. We thank O. Salem of McMaster Immunology Research Centre for Fig. 3i and M. AbdelMoety of American University of the Middle East for the chemical structure in Fig. 3i. This work was supported by ERC AdG 268515, DFG SFB873 and NIH GM124434 to B.A.E., and by the Helmholtz Zukunftsthema 'Aging and Metabolic Programming' (AMPro) to A.A.T.

Author contributions S.M.H.A. performed and analysed all experiments except Extended Data Figs. 8m, 9a–d (G.O.P.-S.) and Extended Data Fig. 6a, e–q (C.P.). J.A.M. contributed to Extended Data Fig. 8a, n and Fig 2o. The conception and design of experiments was by S.M.H.A., B.A.E., J.A.M. and A.A.T. Image-processing methods were designed by D.K. S.M.H.A., A.A.T. and B.A.E. wrote the manuscript.

Competing interests The authors declare no competing interests.

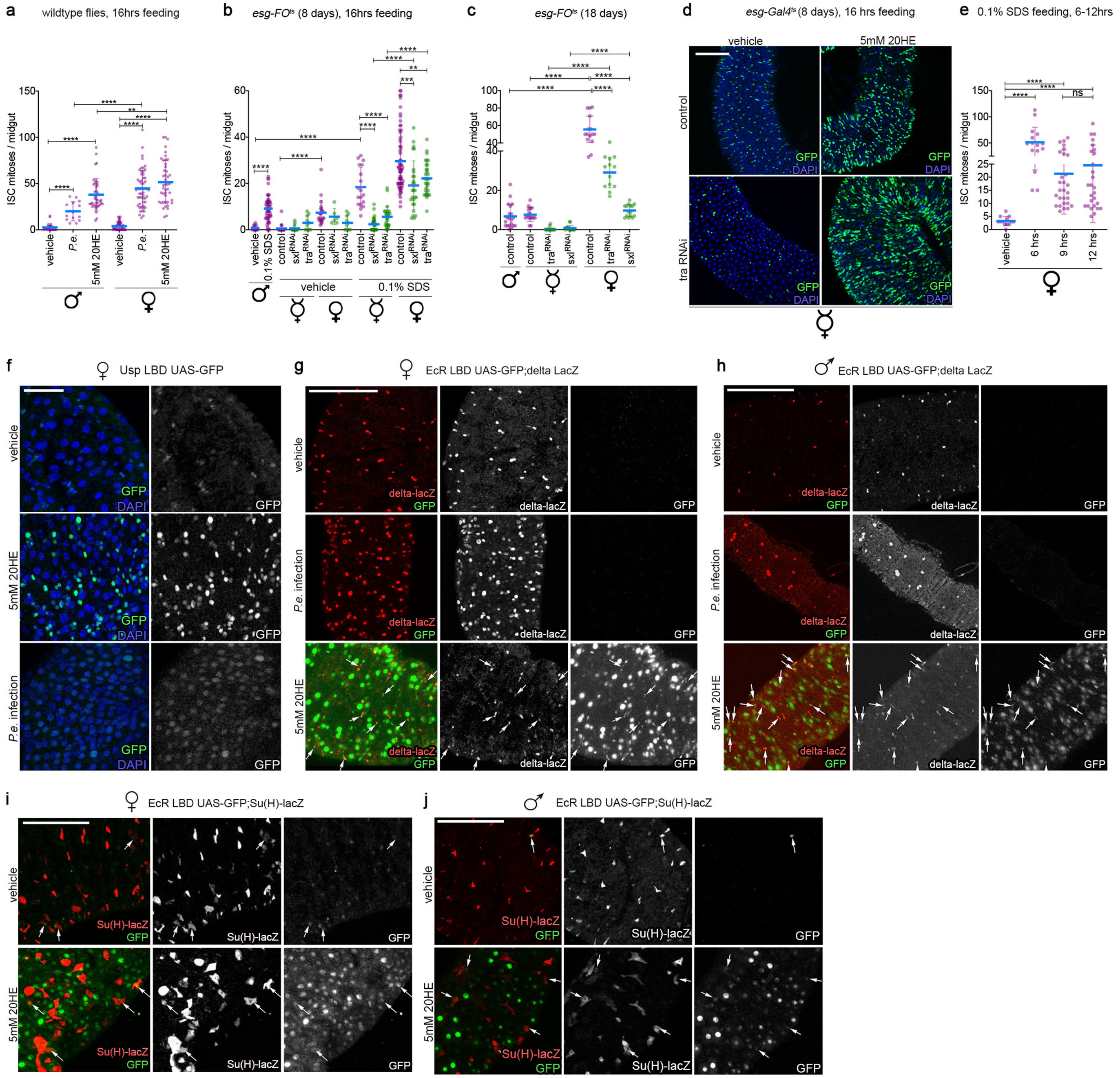
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2462-y.

Correspondence and requests for materials should be addressed to A.A.T. or B.A.E.

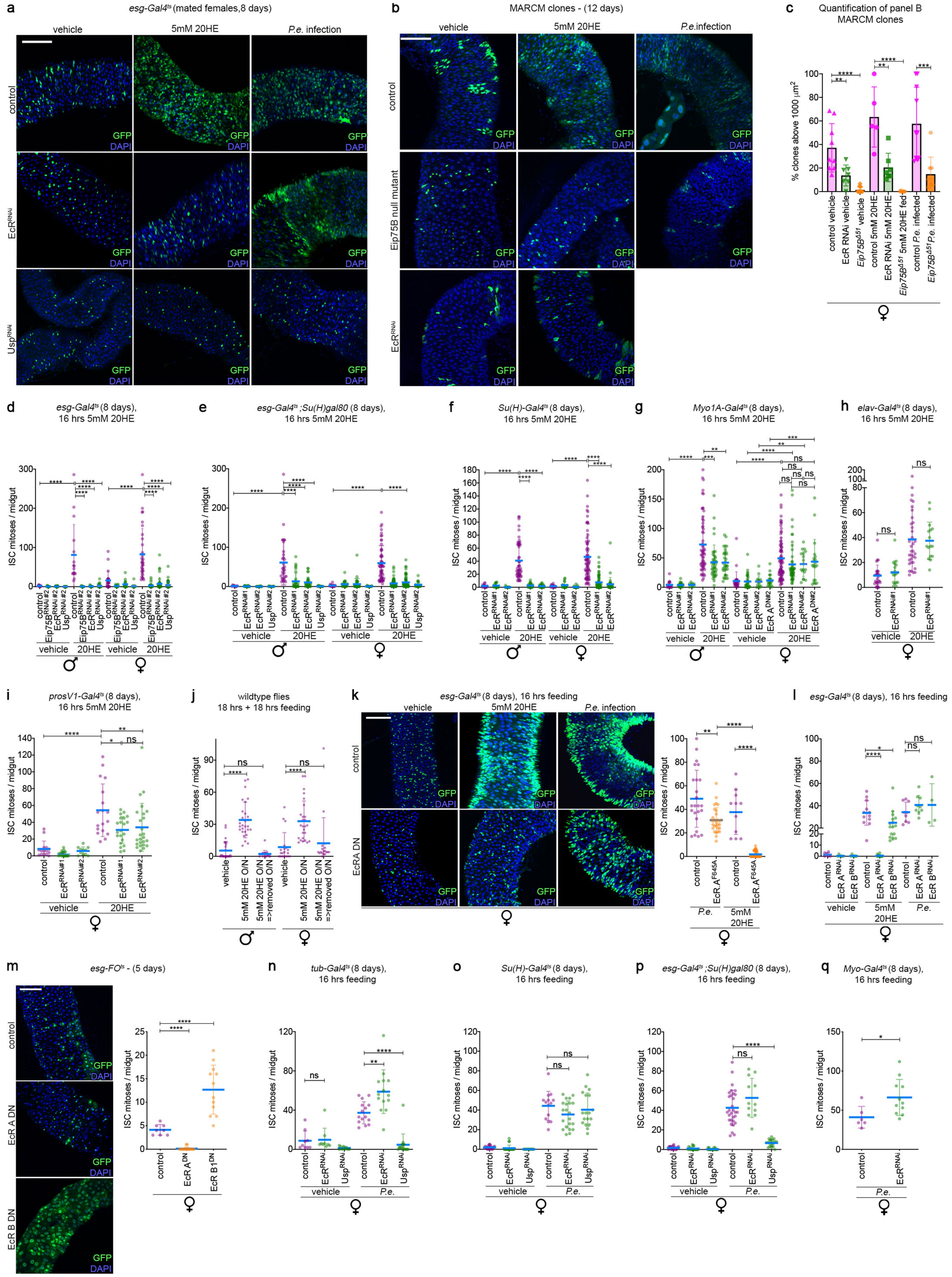
Peer review information *Nature* thanks Henri Jasper, Marc Tatar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



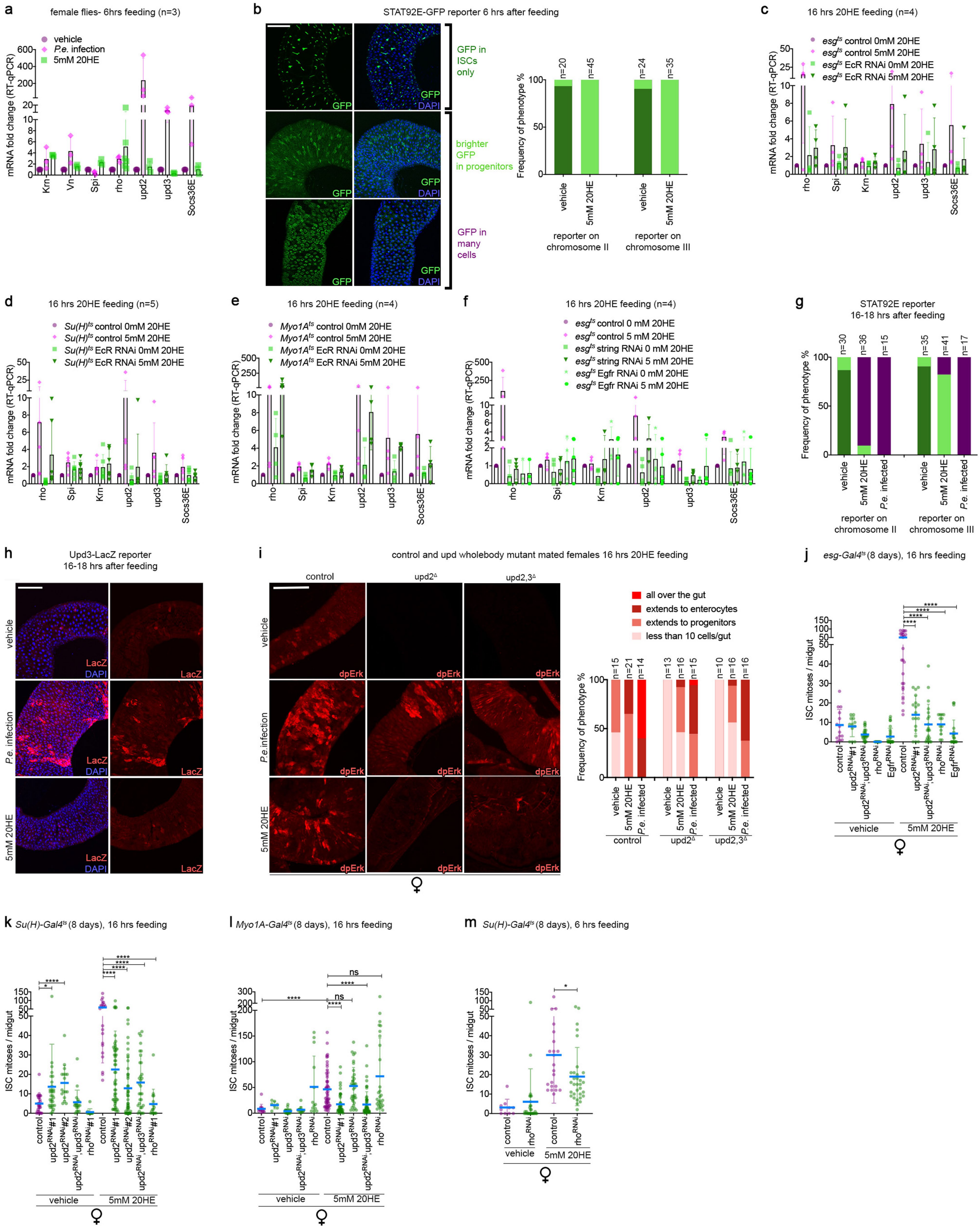
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | 20HE feeding promotes sexually dimorphic ISC mitotic activity. a, Male ISCs do not divide strongly in response to infection elicited by pathogenic bacteria, but divide to a similar extent as mated female ISCs in response to 20HE feeding, quantified by counting the number of dividing ISCs per midgut using pH3 staining (also termed the mitotic index) in males and mated females after 16–18 h treatment with 5 mM 20HE or pathogenic *P.e.* infection. Males are fully and equally competent to respond to 20HE treatment as mated females. **b**, Mating boosts the mitotic divisions of ISCs. Feeding 0.1% SDS for 16 h to virgin females induces ISCs mitoses and this is inhibited by masculinizing ISC clones using *sxl* or *tra* RNAi. Mating increases the ISC mitotic responses to SDS feeding and restores the ability to masculinized ISCs to divide to stress. **c**, Mating induces basal ISC mitoses in both female (control) ISCs and in masculinized ISC clones with *tra* or *sxl* depletion. **d**, 20HE feeding leads to the proliferation and expansion of both control ISCs and ISCs of tra^{RNAi} masculinized progenitors. Representative images are shown 16 h after 5 mM 20HE feeding. This experiment was repeated three times with similar results. Quantification is shown in Fig. 1a. **e**, Quantification of ISC division at different time points (6, 9 and 12 h) after feeding 0.1% SDS to mated females. **f**–**j**, Males or mated females of the genotypes *Gal4.DBD-Usp.LBD>GFP* (*Gal4-Usp>GFP*) (**f**) or *Gal4.DBD-EcR.LBD>GFP* (*Gal4-EcR>GFP*) (**g**–**j**) were heat-shocked for 30 min to induce expression of the ligand sensor system, and then either infected with *P.e.* or fed with 5 mM 20HE or vehicle and dissected 18–20 h later. These GFP ligand traps express GFP under the control of heat-inducible promoter and mark cells with active 20HE signalling. When fed with vehicle, both *Gal4-EcR>GFP* and *Gal4-Usp>GFP* flies were expressed in a few cells in the R4 region posterior midgut (image shown) and in many more in the anterior midgut (image not shown). White arrows indicate cells that are doubly positive for delta or Su(H) *lacZ* markers. Feeding of 5 mM 20HE caused a strong increase in GFP expression in the posterior midgut, indicating an upregulation in the activity of both reporters. GFP was expressed in many delta^{+} cells (**g**, **h**) and much fewer Su(H)$^{+}$ cells (**i**, **j**) of both males and females after 5 mM 20HE feeding. Most of the remaining positive cells are enterocytes. After 20 h of *P.e.* infection, the GFP signal disappears from males and females guts, indicating that EcR is not involved in infection-induced stress response (**g**, **h**). However, the Usp reporter was still active in many gut cells as a consequence of *P.e.* infection (**f**). The Usp reporter was also positive in many cell doublets and bigger cells of the midgut. These reporter data suggest that EcR and Usp are both activated by exogenous 20HE feeding, but they act differently in response to infection. Representative images are shown. This experiment was repeated five times with similar results. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from at least three independent biological replicates. Data are mean and s.d. $n \geq 10$ are plotted for each genotype in each scatter plot. $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the Source Data. Scale bars, 50 µm (**f**) or 100 µm (**d**, **g**–**j**). The overnight standard period of feeding the flies was 16–20 h.



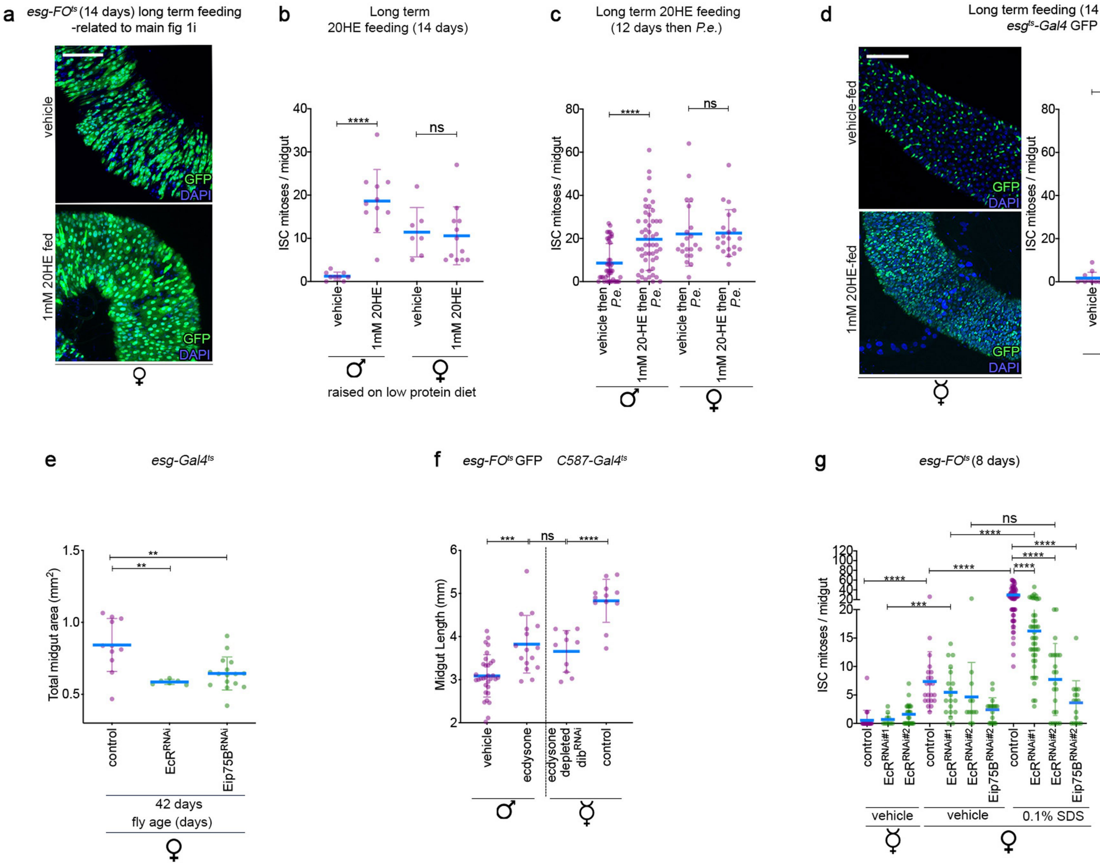
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | The second mitotic wave of 20HE requires EcR–Usp in progenitors whereas EcR is dispensable to ISCs in their response to *P.e.* infection. a, Representative images of samples from Fig. 1c, h. Both EcR and Usp are required in progenitors for the mitoses induced 16 h after 20HE feeding, whereas only Usp is cell-autonomously required by the ISCs for *P.e.*-induced mitoses. Shown are images of progenitor accumulation in mated females after 20HE feeding or *P.e.* infection. **b**, ISCs depleted of EcR or its downstream target Eip75B are unable to form clones in response to 20HE feeding. Eip75B-null mutant clones also fail to regenerate the epithelium after *P.e.* infection. EcR-depleted or Eip75B-null mutant clones were generated by MARCM and analysed 12 days after clonal induction followed by 5 mM 20HE feeding or *P.e.* infection for 16–18 h. Vehicle-fed control clones were multicellular and spread throughout the epithelium, whereas EcR-depleted clones were considerably smaller, mostly between two and four cells, and rarely up to ten small cells per clone. Eip75B-null mutant clones remained mostly single ISC clones. After 16 h of 20HE feeding, the epithelium is populated with newly formed cells within the control clones; however, both EcR- and Eip75B-depleted clones remained unable to divide, indicating the ISC cell-autonomous requirement of EcR and Eip75B for ISC mitoses both basally and in response to exogenously fed 20HE. Similarly, after *P.e.* infection, GFP^+ cells expanded in control clones, whereas Eip75B-null mutant clones were considerably smaller. **c**, Quantification of data in **b** by a macro designed to assess clonal sizes/maximum *Z* projection (Methods, Supplementary Data 2). **d**, Both EcR and Usp are required in gut progenitor cells for the 20HE-induced-mitotic response as shown by the reduced ISC mitotic activity 16 h after feeding 5 mM 20HE to flies with progenitor-specific depletion of EcR or Usp in males and mated females. Results shown are for a second RNAi line to complement the results in Fig. 1c. **e**, EcR or Usp depletion in ISCs abolishes ISC mitoses 16 h after feeding 5 mM 20HE to males and mated females. Results shown are for two different RNAi lines. **f**, EcR is required in EBs for the second wave of ISC mitoses induced 16 h after feeding 5 mM 20HE to males and mated females. Results shown are for two different RNAi lines. This experiment indicates that in contrast to the first wave (Fig. 1e), EcR is required non-cell autonomously in EBs for 20HE induced ISC divisions. **g**, EcR is non-autonomously required in ECs for maximal induction of ISC mitoses in response to 20HE. The *Myo1A-Gal4*^{ts} driver (*Myo1A-Gal4 tub-Gal80*^{ts}) activates UAS target gene expression specifically in ECs. Results shown are for two different RNAi lines for both males and females, and for a dominant-negative isoform of EcR (EcR-A$^{DN\#2}$) in females. **h**, EcR in the nervous system is not required for intestinal 20HE-stem-cell induced mitoses. EcR depletion was induced using *elav-Gal4 tub-Gal80*^{ts}, a pan-neuronal driver for the adult central nervous system. Sixteen hours after 5 mM 20HE feeding, ISCs mitoses were scored and midguts with EcR depletion in the CNS did not exhibit a change in their division rates in comparison to control females. **i**, EcR in enteroendocrine cells has a minimal role in 20HE-induced ISC mitoses of the midgut. Slightly compromised mitotic indexes in 20HE-fed mated females after enteroendocrine cell-specific depletion of EcR in EEs using the enteroendocrine cell-specific *prosV1-Gal4 tub-Gal80*^{ts} driver indicate that EcR in enteroendocrine cells is dispensable to the 20HE induced ISC mitoses. Results shown are for two different RNAi lines. **j**, 20HE only transiently induces ISC mitoses, quantified by mitotic indices of male and female wild-type flies subjected to two-day of the indicated treatment regimes. ISC proliferation is restored to basal levels after 5 mM 20HE was withdrawn, which suggests that the actions of 20HE are not detrimental. Male and female flies were fed vehicle or 20HE in different successions such that flies were exposed for 20 h to the first treatment, then for another 24 h to the second treatment. ISC mitoses returned to basal levels after 16–20 h treatment with 20HE then vehicle. **k**, Expression of an EcR-A dominant-negative isoform inhibits the ISC proliferative response to 5 mM 20HE but not to enteric infection. Left, images of progenitors marked with *esg-Gal4* after *P.e.* infection or 5 mM 20HE feeding, indicative of ISC proliferation in control mated females. Right, mitotic counts. **l**, 20HE signals mostly through isoform EcR-A to mediate ISC proliferation. Progenitor-specific expression of *EcR-A*^{RNAi} and *EcR-B*^{RNAi} shows that EcR-A, more than EcR-B, is required in ISCs for their mitotic response 16–20 h after feeding of 20HE. Knockdown of neither EcR-A nor EcR-B had an effect on the *P.e.*-induced ISC mitoses. **m**, EcR isoform A is much more important than isoform B for driving the intestinal hyperplasia, as shown in images of posterior midguts of mated females expressing different EcR dominant-negative isoforms. Left, images of clonal expansion under basal conditions at 5 days after induction of expression of different EcR dominant-negative isoforms in mated female midguts. Right, ISC mitotic counts. **n**–**q**, EcR in ISCs or other differentiated cells is not required for the *P.e.*-induced mitotic response of ISCs, whereas Usp is cell-autonomously required by ISCs to proliferate in response to *P.e.* infection. Quantification of the mitotic indexes of ISCs after *P.e.* infection in mated females in which EcR or Usp was depleted: constitutively in all cells using the *tub-gal4*^{ts} driver (**n**), in EBs (**o**), in ISCs (**p**) or in ECs (**q**). Collectively, these experiments indicate a functional bifurcation of EcR and Usp, in which Usp is essential in ISCs for the *P.e.*-induced ISC response. RNAi was induced in progenitors of mated females for 8 days before 16–20 h of *P.e.* infection or 20HE feeding. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from three independent biological replicates. Data are mean ± s.d. $n \geq 10$ are plotted for each genotype in each scatter plot. $^*P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the Source Data. Representative images are shown from experiments that were repeated three times. Scale bars, 100 µm. The overnight standard period of feeding the flies was 16–20 h.

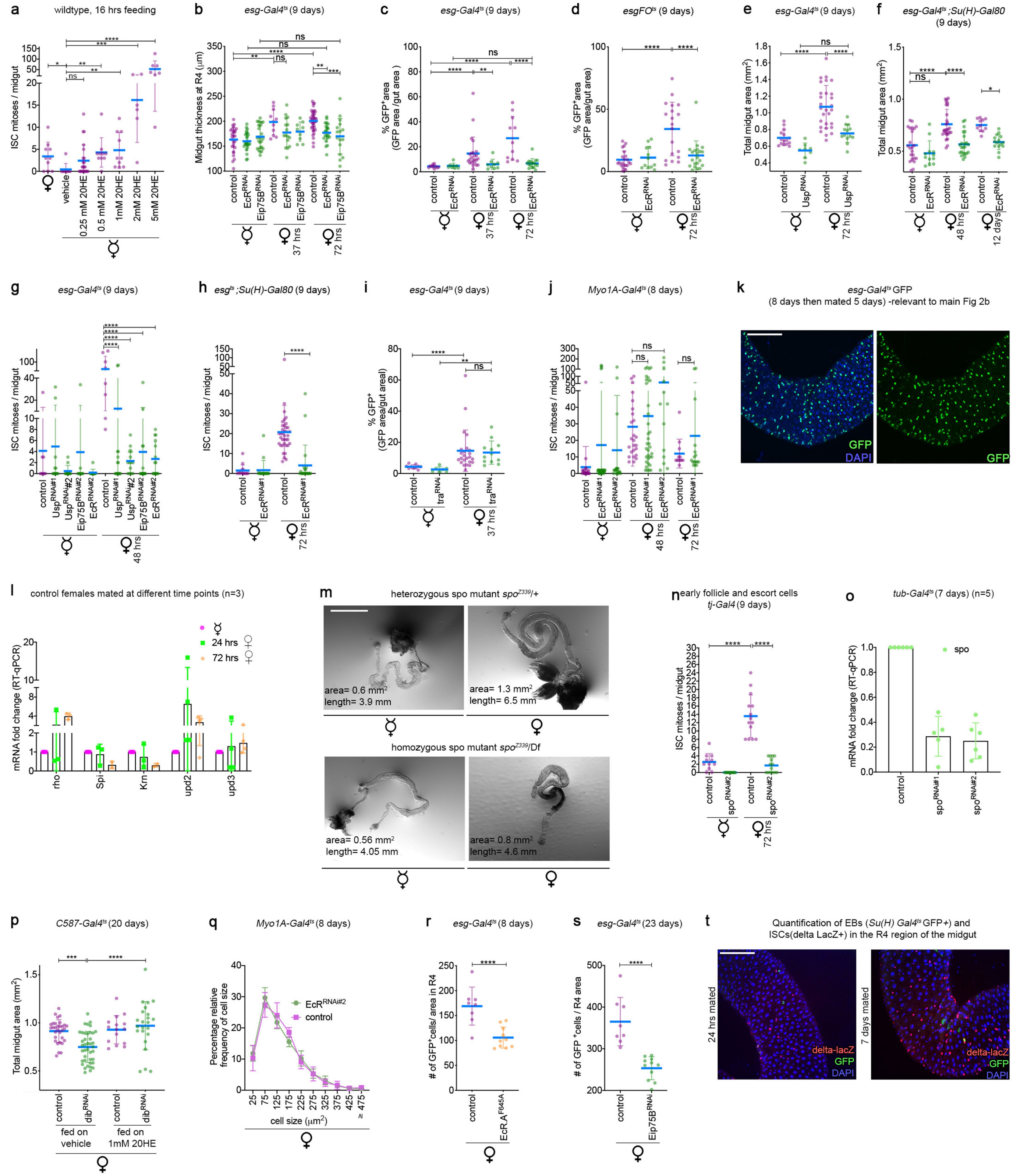


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | The second mitotic wave of 20HE regulates Jak–Stat signalling and requires Egfr signalling in the midgut progenitors. **a**, Components of Egf signalling but not the Jak–Stat pathway are transcriptionally induced 6 h after 20HE feeding. mRNA levels of Egf ligands such as *keren*, *spitz* and their cleaving protease *rho* are transcriptionally induced whereas unpaired cytokines *upd2*, *upd3*, and Jak-Stat target *Socs36E* are not induced 6 h after 20HE (light green bars) relative to vehicle-fed control females (dark pink bars). By contrast, *P.e.* infection causes a strong induction of Jak-Stat signalling components *upd2*, *upd3*, *Socs36E* as well as a milder upregulation of Egf signalling components *keren*, *vein* and *rho* (light pink bars). Mated female midguts of wildtype flies were fed with vehicle, *P.e.* or 5 mM 20HE for 6 h then expression levels in guts were determined by RT–qPCR. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n = 3$). **b**, Left, representative images of three categories of activity for the phenotypes of STAT92E-GFP reporters on chromosome II or III. The frequency of phenotype was quantified (right; and in **g**) in reference to phenotypes observed in the R4 region. Dark green text/bars denote no activation of the reporter. Bright green text/bars denote a mild activation pattern. Purple text/bars denote the strongest activation pattern. 5–7-day-old mated females were used for the experiment. Right, under homeostatic conditions, the reporter expresses GFP only in ISCs (dark green bar). At 6 h after 20HE feeding, GFP is localized in midgut progenitors all over the gut (bright green bar). 18% of the guts that express the reporter on chromosome II show a slight accumulation of GFP in other cells after 20HE feeding, but the GFP signal was not as strong as in the category 'GFP in many cells'. **c**–**e**, EcR is required in midgut progenitors (**c**) and EBs (**d**) but not ECs (**e**) for transcriptional induction of *rho*, *upd2* and *upd3* during the second mitotic wave in response to 20HE feeding. By contrast, induction of *spitz* and *keren* are unchanged relative to 20HE fed controls. qRT–PCR was performed on midguts from mated females 8 days after RNAi induction at 29 °C followed by feeding with vehicle or 5 mM 20HE for 16 h. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n \geq 3$). **f**, ISCs need to proliferate in order for *rho*, *upd2* and *upd3* to be induced during the second mitotic wave after 20HE feeding. Egf and Jak–Stat signalling are transcriptionally induced 16 h after 20HE feeding. Control midguts have a transcriptional induction of *rho*, *upd2* and *Socs36E* and to a lesser extent *upd3* mRNA levels (vehicle denoted as purple versus control 20HE-fed denoted as pink bars). Cell cycle arrest via string depletion or reduced Egfr signalling in midgut progenitors halts the upregulation of 20HE-induced *rho*, *upd2*, *Socs36E* and *upd3*. These data suggest that ISC division is cell autonomously controlled and this event is an initial requirement for the non-cell autonomous induction of promitotic factors to promote later ISC divisions. mRNA induction of *spitz* and *keren* is slightly decreased in string-depleted progenitors but are slightly higher in Egfr-depleted progenitors relative to 20HE fed controls. Mated female midguts of wild-type flies, *string* or Egfr-depleted progenitors for 8 days at 29 °C were fed with vehicle or 5 mM 20HE for 16 h then expression levels were determined by RT–qPCR. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n \geq 3$). **g**, 20HE feeding induces activity of a Jak–Stat reporter more mildly than *P.e.* infection. Frequency of phenotype occurrence is analysed based on the categories of activity in **b**. Under homeostatic conditions, the reporter expresses GFP only in ISCs (dark green bar). Sixteen hours after 20HE feeding, most midguts of the reporter on chromosome II have GFP localized in many midgut cells including polyploid ECs (purple bar). However, most midguts of the reporter on chromosome III have GFP localized in the midgut progenitors (bright green bar). By contrast, *P.e.*-infected midguts of the reporters on either chromosome showed a strong uniform activation pattern in all midgut cells of the R4 region. 5–7-day-old mated females were used for the experiment. **h**, The upd3-lacZ reporter is not activated by 20HE feeding. Images of the R4 region of the midgut showing basal expression of the upd3 reporter in vehicle-fed flies relative to strong activation of the reporter after *P.e.* infection. By contrast, 16 h of 20HE feeding did not appreciably activate the upd3 reporter. These data indicate that 20HE does not primarily activate upd3 to promote ISC mitoses in the midgut. 5–7-day-old mated females were used for the experiment. All images were acquired at the same settings and the intensities of activation are accurately represented. **i**, Left, representative images of Erk activity, assayed as dpErk showing the most prevalent phenotype for each condition. Right, quantifications of the prevalence of each phenotype are shown. Under non-stressed conditions, dpErk is present either in very few ECs per gut, or in progenitor cells and very few ECs. After enteric infection, there is a strong upregulation of dpErk mainly in ECs. Although 20HE feeding also induces dpErk in midguts, the pattern is distinct from the one caused by enteric infection. After 20HE feeding, dpErk in mainly visible in progenitors and young ECs, and the signal is often localized to small patches of cells. By contrast, *P.e.* infection induces strong dpErk broadly throughout the gut. dpErk is absent in non-stressed *upd2* or *upd2,3* mutants. Enteric infection induces dpErk also in *upd2* or *upd2,3* mutants, albeit to a lower level than wild-type flies. By contrast, *upd2* or *upd2,3* mutants show very little or no dpErk after 20HE feeding. 5–8-day-old mated females were used for the experiment. **j**, Upd2, Egfr and rho are required in gut progenitors for the second wave of mitoses induced by 20HE as shown by the diminished ISC mitoses 16 h after feeding 5 mM 20HE to mated females with progenitor-specific depletion of Upd2, Upd2+Upd3, Egfr or rho. **k**, Upd2 and rho are required in EBs for the second wave of mitoses induced by 20HE as shown by the diminished ISC mitoses 16 h after feeding 5 mM 20HE to mated females with EB-specific depletion of Upd2, Upd2 and Upd3 or rho. Results shown are for two different RNAi lines for Upd2. **l**, Upd2 but not Upd3 or rho is required in ECs for the second wave of mitoses induced by 20HE as shown by the diminished ISC mitoses 16 h after feeding 5 mM 20HE to mated females with enterocyte-specific depletion of Upd2, Upd2 and Upd3 or rho. **m**, Rho is partly required in EBs for the optimal ISC mitoses during the first mitotic wave in response to 6 h of 20HE feeding. ISCs were still able to divide at 6 h of 20HE feeding after rho depletion in EBs albeit at lower but non-significant levels relative to control flies. This result indicates that ISCs, with their intrinsic EGF signalling retain the ability to divide in response to 20HE in a cell-autonomous fashion. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from three independent biological replicates except for the qPCRs in which the *n* numbers are indicated. $N \geq 10$ are plotted for each genotype in the remaining scatter plots. Data are mean ± s.d. $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the Source Data. Representative images are shown from experiments that were repeated at least three. Scale bars, 100 µm. The overnight standard period of feeding the flies was 16–20 h.

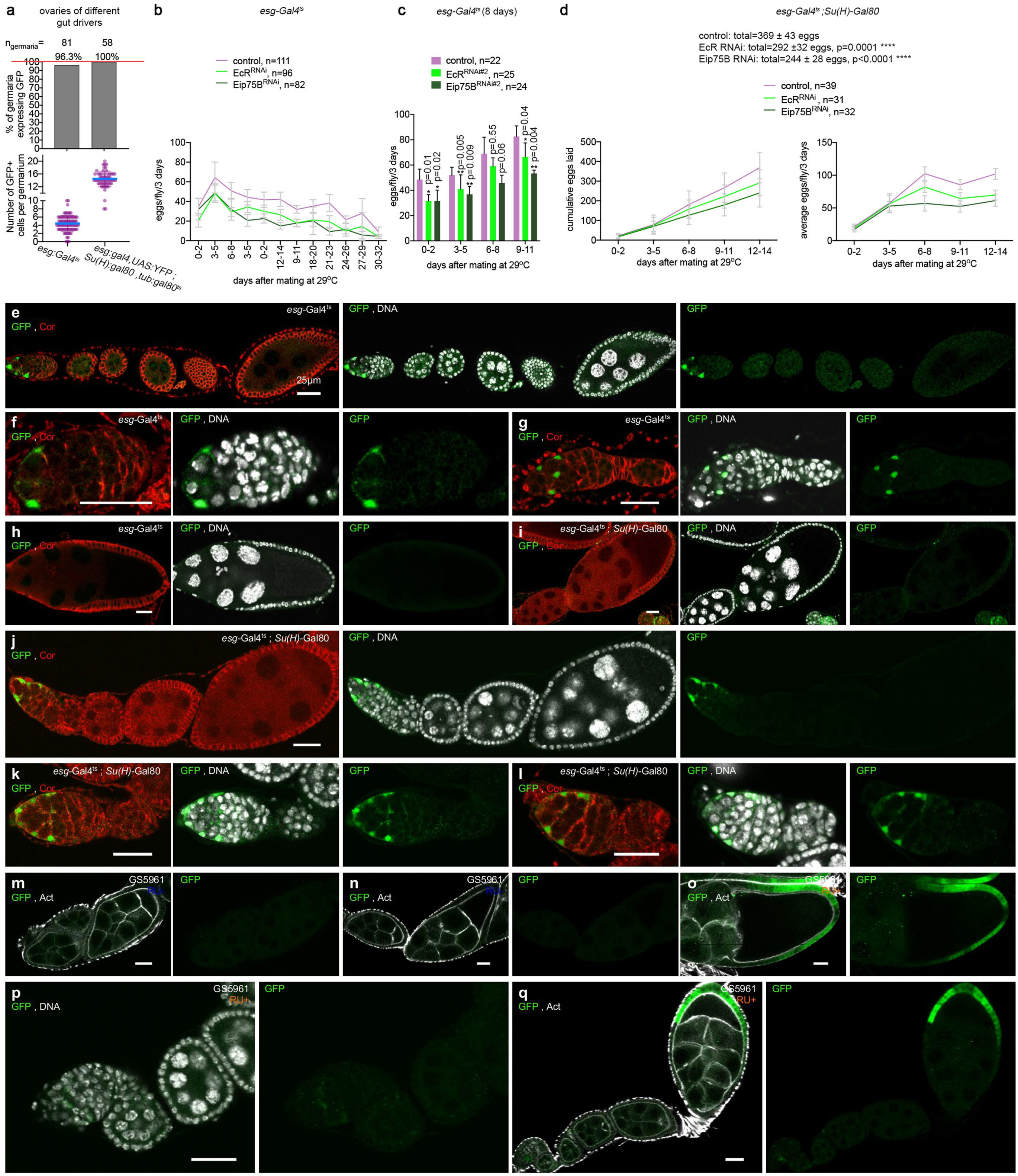


Extended Data Fig. 4 | Long-term 20HE feeding promotes sexually dimorphic ISC division and gut growth. a, 1 mM 20HE feeding does not obviously increase epithelial turnover in females. Representative images are shown and are relevant to Fig. 1i. **b**, 20HE feeding causes male-specific midgut growth also on a low-protein diet, quantified by counting mitotic indexes of males and females raised on 20HE-laced low-yeast sucrose solution or sucrose-yeast solution as vehicle. 20HE- or vehicle-fed female ISCs did not differ in their mitotic counts. However, 20HE-fed males had a strong increase in their mitotic index compared to vehicle-fed males. **c**, 20HE feeding enhances ISC mitotic activity in *P.e.*-infected males, altering their behaviour to resemble *P.e.*-induced ISC division in females, assayed by mitotic counts of males and females. Flies were raised on 20HE or vehicle-supplemented food for 12 days then the treatment was withdrawn overnight followed by *P.e.* infection for 20 h. Male ISCs that were 20HE-fed were able to respond to *P.e.* infection at similar rates to the age-controlled females fed on 20HE or vehicle. **d**, 20HE-fed virgins undergo epithelial turnover much faster than age-controlled virgins, which have infrequently dividing ISCs. Representative images (left) and quantification (right) of mitotic counts from control virgin flies 14 days after 20HE feeding. Both the frequency of dividing ISCs and progenitor cells of 20HE fed virgins resemble the behaviour of mated females. **e**, Eip75 and EcR are required in midgut progenitors to maintain proper midgut size, quantified as midgut areas in images of guts from mated females with progenitor-specific depletion of EcR or Eip75B aged for 42 days. **f**, Quantification of midgut lengths of control males, 20HE-fed males, control virgin females, or virgin females depleted of ecdysone via ovary-specific knockdown of *dib*^{RNAi}, shows the plasticity of male and female midgut growth to 20HE levels. 20HE-fed males have increased midgut length in contrast to *dib*^{RNAi} female virgins, with decreased 20HE levels and notably shorter guts. In both cases, there was a one-third gain or loss in midgut length in comparison to a control male or virgin female, respectively. **g**, Ecdysone signalling via EcR and Eip75B is required in ISC clones of mated females for maximal proliferation in response to SDS. ISC mitotic counts of virgin females are minimal under basal conditions. After SDS feeding, control ISC clones divide to regenerate the epithelium but EcR- or Eip75B-depleted ISC clones are significantly impaired in their ability to divide. RNAi was induced in ISC clones for 8 days before 16–18 h of 0.1% SDS feeding. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from three independent biological replicates. $N \geq 10$ are plotted for each genotype in the remaining scatter plots. Data are mean ± s.d. $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P < 0.0001$, Mann–Whitney test with two-tailed distribution (all panels except **f**) or ordinary ANOVA test followed by Bonferroni's multiple comparisons test (**f**). Exact *n* numbers and *P* values are in the online Source Data. Long-term 20HE feeding indicates that 1 mM 20HE was fed to the flies for 12 (**c**) or 14 days (**a**, **b**, **d**). Representative images are shown from experiments that were repeated three independent times. Scale bars, 100 μm.



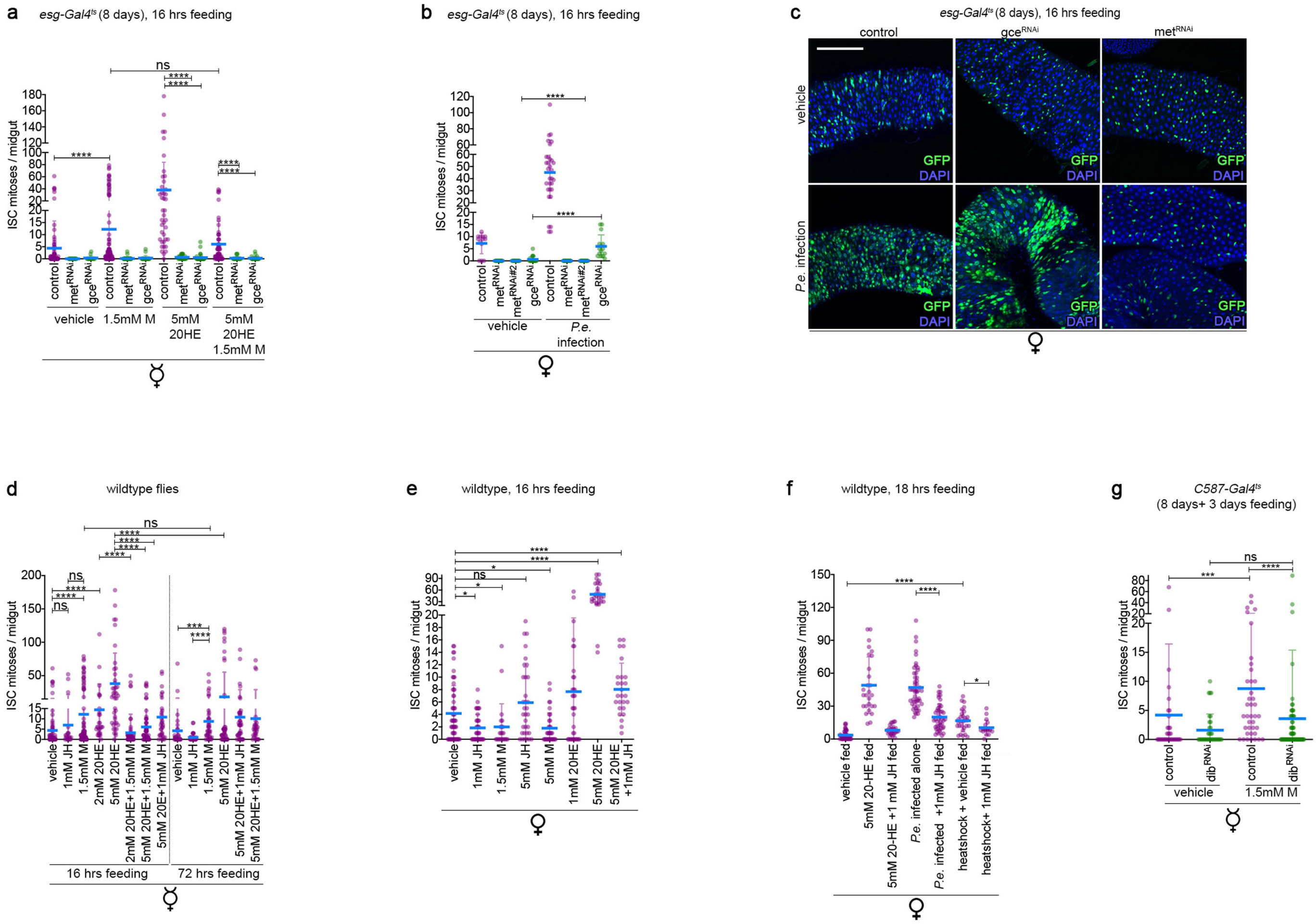
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Mating requires ecdysteroidogenic enzymes from the early ovarian follicles and escort cells to induce ISC divisions in the gut through EcR–Usp, which causes increased stem-cell number and subsequent gut growth. a, 20HE induces ISC mitoses in a dose-dependent manner in ISCs of virgin females. Virgin females were fed with different doses of 20HE and their mitotic indexes were assessed after 16–18 h of feeding. At 0.25–1.00 mM 20HE, ISCs divide similar to basal levels in mated females. At 2 mM 20HE, ISCs mildly divide (3–4 times higher than divisions induced by 1 mM 20HE). At 5 mM 20HE, ISCs divide at 10–11 times higher that divisions induced by 1 mM 20HE. **b**, The increase in width of the R4 region in response to mating in females requires EcR and Eip75B in progenitors. **c**, EcR is required in intestinal progenitors for their accumulation upon mating, shown by quantification of the GFP$^+$ labelled areas of progenitors in the midgut after progenitor-specific depletion of EcR ± mating at early and later time points after mating. **d**, EcR-depleted ISC clones are unable to divide in response to mating, as quantified by the GFP$^+$ clonal area in EcR-depleted ISC-derived clones and age-matched control clones. ISC-derived clones in control females have GFP$^+$-labelled ISCs and all their subsequent progeny stably express GFP as well. **e**, Usp is required in progenitors for the mating-induced midgut growth as shown by quantification of midgut areas in females with Usp-depleted progenitors ± mating. **f**, EcR is cell-autonomously required in ISCs for mating-induced midgut growth, shown by quantification of midgut areas in females with EcR-depleted ISCs ± mating. After the first mating, control female midgut initially grows and midgut growth persists in flies that are raised repeatedly mated. This midgut growth requires EcR functions in ISCs. **g**, Ecdysone signalling via EcR, Usp and Eip75B are required in midgut progenitors for the mating-induced mitotic response, as shown by the reduced ISC mitoses upon 48 h mating in female midguts with progenitor-specific depletion of EcR, Usp or Eip75B. Virgins were left to mate for 48 h before dissection, then mitotic counts were assessed. Results shown are for a second RNAi line to complement the results in Fig. 2. **h**, EcR is cell-autonomously required in ISCs for mating-induced ISC mitoses shown by mitotic counts of midgut in females with EcR-depleted ISCs 72 h after mating. Results shown are for a second independent RNAi to complement Fig. 2f. **i**, Masculinized *tra*^{RNAi} progenitors undergo mating-induced expansion of GFP$^+$ progenitors similar to controls, indicating that the mating effects on progenitors are independent of the sex determination pathway, quantified as GFP$^+$ area of progenitors in the R4 region. Virgins typically have GFP-marked single cells (ISCs) or few pairs (ISC-EB). Shortly after mating, the ISC cells divide and the resulting progeny are transiently marked with GFP, but then turn off GFP expression as they differentiate. **j**, EcR is not required in ECs for mating-induced ISC mitoses. 48 h to 72 h after mating, ISCs of EcR depleted ECs midguts divide at similar rates to control midguts indicating that EcR in ECs is dispensable to mating-induced ISC mitoses. Results shown are for two different RNAi lines. **k**, Representative confocal image of GFP-expressing progenitors using *esg*^{ts} in females 5 days after mating. Flies were raised as virgins and were aged for 8 days (similar to conditions in Fig. 2b), and then mated for 5 days. Females were always mated to males with no genetic manipulations. Equal number of males and females were allowed to mate (a ratio of 1:1). Image is acquired in the R4 region. This suggests that the strong mitotic effect of mating is transient. Scale bars, 100 µm. **l**, *Rho* and *upd2* are transcriptionally upregulated in female midguts 24 h (green) or 72 h (orange) after mating relative to virgins (pink). 5–7-day-old control virgins were mated for 24 or 72 h, then mRNA expression levels were determined by RT–qPCR. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n = 4$). **m**, Representative images of whole-body *spo* mutants that are either heterozygous and hence viable with no growth or egg-laying defects (top) or sterile, homozygous *spo* mutants rescued to adulthood with by a pulse of 20HE given to dechorionated embryos (bottom). Images are complementary to Fig. 2i. Scale bars, 1 mm. **n**, RNAi-mediated depletion of *spo* in ovaries blunts ISC mitoses in response to mating. The traffic jam (*tj-Gal4*) driver that is expressed in somatic gonadal cells was used for spo depletion. Flies were raised as virgins then mated for 72 h. **o**, *spo*^{RNAi} depletes the *spo* gene efficiently. Constitutive driver *tub*^{ts} was used to deplete *spo* in mated females for 8 days, and then mRNA expression levels were determined by RT–qPCR. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n = 4$). **p**, Ovary-derived ecdysone is required for the proper size of the midgut, shown by quantification of midgut areas in mated female midguts depleted of 20HE-synthesizing enzyme Dib in the ovary. The *C587*^{ts} driver, which is expressed in escort cells and immature follicle cells of the ovary, is used to induce ecdysteroidogenic enzymes depletion. Decreased midgut area in mated females with reduced 20HE levels is completely rescued by raising females on exogenous 1 mM 20HE. *Dib*^{RNAi} was previously validated[35]. **q**, Depletion of EcR in midgut ECs does not significantly decrease their size 8 days after mating. Cells of the midgut were stained with CellMask, a plasma membrane stain, and a custom macro (Supplementary Data 1) was used to segment the cells according to size. Shown is a frequency distribution of the different cell sizes. EcR-depleted ECs have a bigger proportion of cells sized 75–175 µm^2 than control midguts. However, the differences in distribution of the cell sizes are statistically non-significant. Data are from $n \leq 5$ stacks of midguts taken at the R4 region. **r**, Basal levels of EcR signalling are required to maintaining the optimal number of progenitors in the midgut as shown by quantification of GFP$^+$ progenitors in mated females expressing dominant-negative EcR-A in comparison to the control. **s**, Basal levels of Eip75B are required for maintenance of ISCs in non-stressed flies, quantified by the number of GFP$^+$ progenitors in mated females after progenitors-specific depletion of Eip75B. A small reduction of progenitor numbers (~25%) suggests that Eip75B is not critical for ISC survival. Note that *y* axis does not go to zero. **t**, Control midguts display an increase of delta$^+$ cells at several time points following mating shown by quantification of delta$^+$ (red) and Su(H)$^+$ (green) cells. At 24 h after mating, most delta$^+$ cells remain singlets, similar to virgins. At 40 h after mating, most delta$^+$ cells expand to become doublets to triplets (Fig. 2k). At 7 days after the first mating most delta$^+$ cells are again singlets; however, their numbers are irreversibly increased relative to virgins. Females were mated to males with no genetic manipulations. Equal number of males and females were allowed to mate (a ratio of 1:1) and females were allowed to mate for 18–20 h after which males were removed, except for the condition 'raised mated for 7 days', in which males were always in the vial with the females. Images are acquired in the R4 region. This suggests that mating induces an initial symmetric increase in the number of ISCs that is irreversible. Representative images for other conditions and quantifications are shown in Fig. 2k. Each dot represents a gut, and the percentage of delta$^+$ or Su(H)$^+$ cells is calculated from absolute number of positive cells relative to total DAPI$^+$ cells. Scale bars, 100 µm. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from three independent biological replicates. $n \geq 10$ are plotted for each genotype in the remaining scatter plots. Data are mean ± s.d. $^*P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P < 0.0001$, ordinary ANOVA test followed by Bonferroni's multiple comparisons test (gut measurements in **b**, **e**, **f**, **p**) or Mann–Whitney test with two-tailed distribution (all other panels). Exact *n* numbers and *P* values are in the online Source Data. Representative images are shown from experiments that were repeated three times.

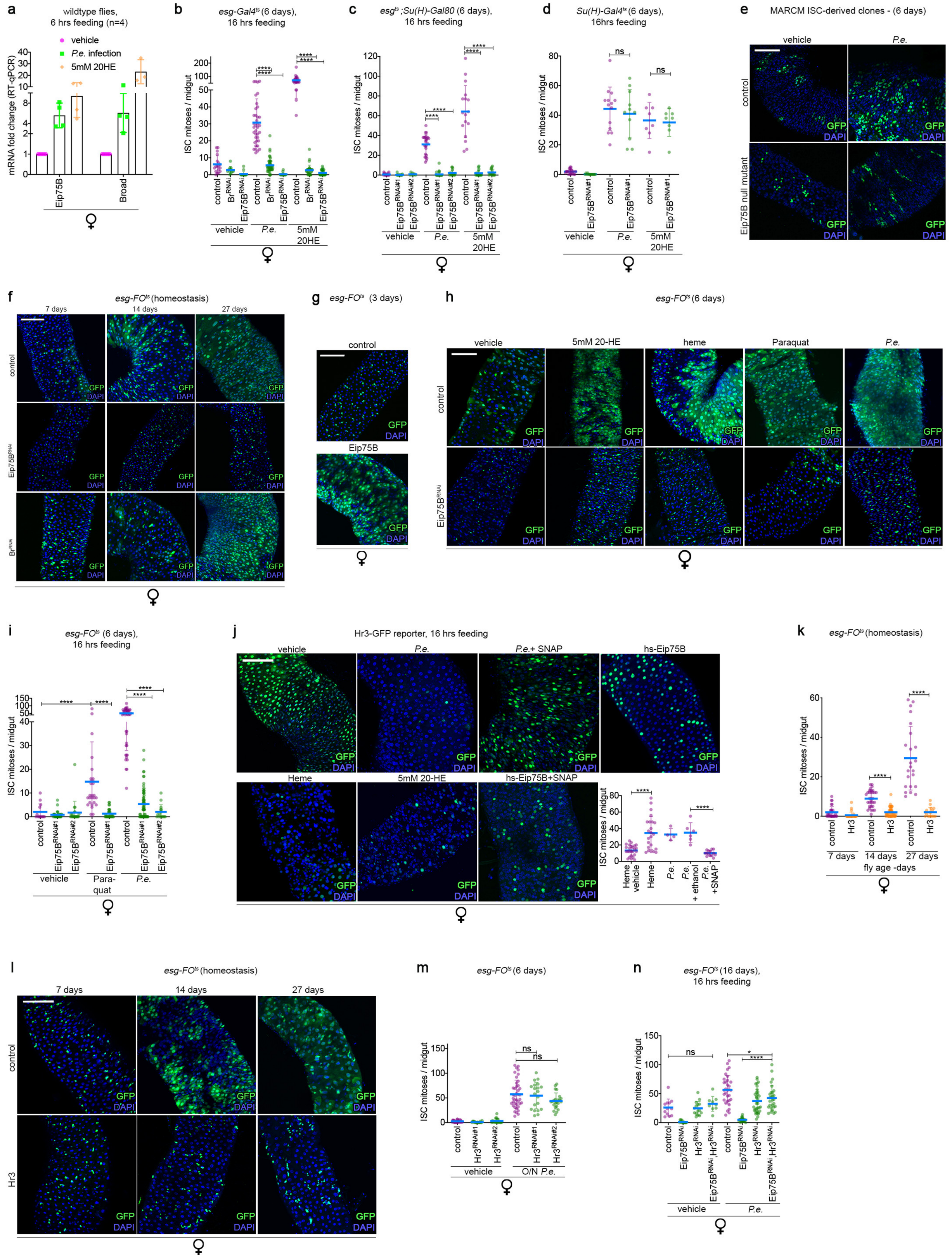


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Ovaries of the *esg-Gal4*^{ts}, *esg*^{ts} *Su*(*H*)*-Gal80* midgut drivers have GFP expression in their germaria in a subset of escort cells. **a**, Top, most midgut drivers express GFP in ovary germaria. The frequency of germaria expressing GFP is displayed in the bar graph. Some ovaries with the *esg*^{ts} driver have no GFP in their germarium while almost all ovaries of the *esg*^{ts} *Su*(*H*)*gal80* driver express GFP. Bottom, the number of GFP^{+} cells per germarium for both midgut drivers *esg*^{ts} or *esg*^{ts} *Su*(*H*)*-Gal80*, which are expressed in midgut progenitors and ISCs respectively. Further examination of *esg*^{ts} driver shows that it is expressed in approximately 4 escort cells, whereas the *esg*^{ts} *Su*(*H*)-*Gal80* driver shows expression in around 14 escort cells. The number of germaria analysed is indicated. Control germaria typically have 45–70 escort cells[36]. **b**, Mated females with EcR- or Eip75B-depleted midguts have reduced reproductive output. This graph is related to Fig. 2p. Average eggs per fly per 3 days are plotted instead of the cumulative sums. Flies that died during the experiment were excluded in the analysis. **c**, Mated females with EcR- or Eip75B-depleted midguts have reduced reproductive output. Flies with control, EcR- or Eip75B-depleted midgut progenitors were raised as virgins for 8 days and then allowed to mate to males with no genetic manipulations at a ratio of 1:1 in populations of 5 females and 5 males. Eggs were collected from the fly vials every day for up to 11 days and the average total eggs per fly every 3 days is plotted. An independent alternative second RNAi is shown to complement data in Fig. 2p. Data are mean and s.d. *P* values were determined by *t*-test with two-tailed distribution assuming unequal variance. **d**, Mated females with EcR- or Eip75B-depleted ISCs have reduced reproductive output. Flies with control, EcR- or Eip75B-depleted midgut ISCs were raised at 18 °C for 2 days maximum and were then shifted to 29 °C and allowed to mate to males with no genetic manipulations at a ratio of 1:1. Flies were pooled together the first night of mating to ensure mating then on the next day, single females were housed with a control male in single vials. Eggs were collected from the fly vials every 48 h for up to 14 days. Flies that died during the experiment were excluded in the analysis. Left, cumulative eggs laid across 14 days ± s.d. Right, the average total eggs per fly every 3 days plotted across 14 days ± confidence intervals. *P* values were determined by *t*-test with two-tailed distribution assuming unequal variance. Exact *n* numbers are in the online Source Data. **e**–**h**, *esg-Gal4*^{ts} and *esg*^{ts} *Su*(*H*)-*Gal80* drive expression in a small number of ovary escort cells. *Drosophila* ovaries are composed of 16 ovarioles. At the anterior tip of every ovariole, the germarium contains the germline stem cells and the somatic stem cells that constantly produce follicles or egg chambers. As the follicles progress to the posterior end of the ovariole, they develop to lead to the formation of a mature egg. Follicle development is divided into 14 stages. In the most anterior part of the germarium (region I) the cap cells and the escort cells constitute the niche required for the maintenance of the GSCs and the proper differentiation of the early germline cyst. We detected expression of the *esg-Gal4*^{ts} and the *esg*^{ts} *Su*(*H*)-*Gal80* drivers within the germarium in a subset of escort cells (**a**). Confocal sections of follicles from stage 2–7 (**e**), stage 9 (**h**) and germaria (**f**, **g**) isolated from *esg-Gal4*^{ts} flies and stained for GFP (green), coracle (red) and DNA (DAPI, grey). No GFP signal was detected in follicles from stage 2 to 9 (**e**, **h**) or in later stages (not shown). However, 96% of germaria showed GFP in a subset of cells in the anterior region I (**f**, **g**). The GFP-expressing cells were located in between the germline cysts and exhibited a triangular shape indicating that they were the escort cells. **i**–**l**, All germaria from *esg*^{ts} *Su*(*H*)-*Gal80*, *UAS-GFP* flies express GFP in escort cells (**a**, **j**, **k**, **l**) and no GFP expression was detected from stage 2 to 9 (**i**, **j**) or in later stages (not shown). **m**–**q**, We detected expression of the Switch *GS5961-Gal4* driver within ovaries in the posterior follicular cells from stage 8 of oogenesis. Confocal section of follicles isolated from *GS5961/UAS-GFP* flies kept on yeast paste only (RU−) or yeast paste supplemented with RU486 (RU+) for 4 days and stained for GFP (green), actin (phalloidin, grey) or DNA (DAPI, grey). In the absence of RU486 induction, no GFP was detected in the ovary (**m**, **n**). After RU486 feeding, no expression was detected in germaria or follicles before stage 7 (**p**, **q**). At stage 7, a subset of the most posterior follicular cells started to express weakly the GFP, this expression was then stronger and spreading to more follicular cells in a posterior to anterior gradient during stage 8 of oogenesis (**q**, most posterior follicle) and maintained later on in most of the posterior follicular cells that cover the oocyte (**o**, stage 10). All pictures are presented with the anterior on the left and the posterior on the right.

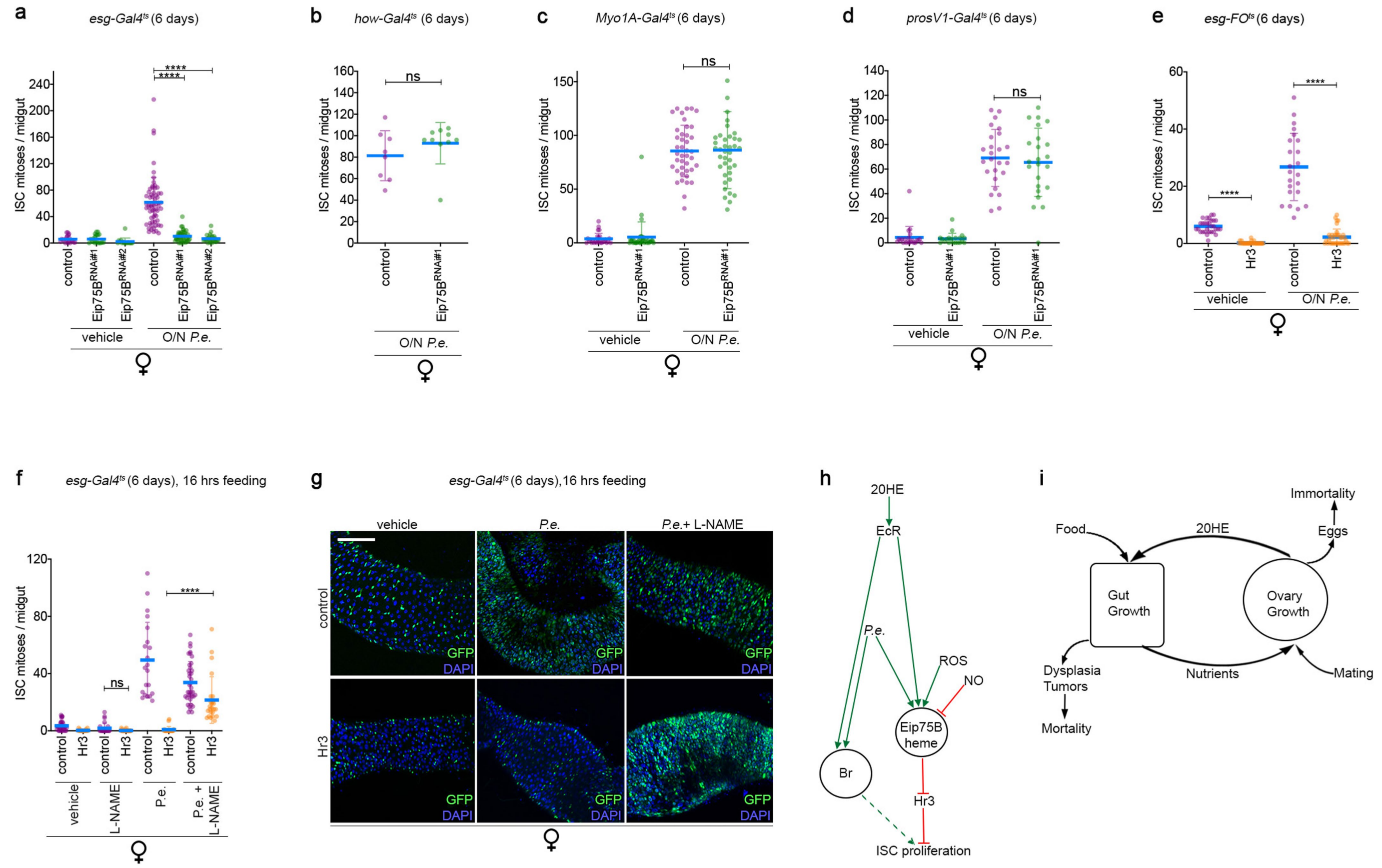


Extended Data Fig. 7 | JH receptors are required for ISC divisions, and exogenously fed JH inhibits ISC mitoses in response to other pro-mitotic stimuli. **a**, JH receptors Met and Gce are required for exogenously fed 20HE to induce ISC mitosis. Virgin females were fed with 1.5 mM methoprene, 5 mM 20HE, or 20HE and methoprene in combination, and their mitotic indexes were assessed after 16–18 h of feeding. Knockdown of Met or Gce in progenitors blunted the proliferative response to all three fed stimuli. Virgins were aged for 8 days at permissive temperature then fed with the different hormone regimes for 16–18 h. **b**, **c**, Met and Gce receptors are required in midgut progenitors of mated females for *P.e.*-induced ISC mitoses. Mated females of indicated genotypes were aged for 8 days at permissive temperature then fed with *P.e.* for 18–20 h. **b**, ISC mitotic counts. **c**, Images of progenitor accumulation after *P.e.* feeding to mated females. **d**, Methoprene induces ISC mitoses in ISCs of virgin females. Virgin females were fed with active JH III ligand (JH), JH agonist methoprene (M), 2 mM or 5 mM 20HE, or the two compounds in combination, and their mitotic indexes were assessed after feeding for 16–18 h (left) or 72 h (right side). After 16–18 h of feeding, the average number of ISC mitoses per midgut was as follows. Vehicle fed: 3.8, 1 mM JH: 6.6, 1.5 mM methoprene: 8, 2 mM 20HE: 14, 5 mM 20HE: 41. A combination feeding of 1.5 mM methoprene with either 2 mM 20HE or 5 mM 20HE blunts mean ISC mitoses to 3.6 or 2.3, respectively. Combination feeding of 1 mM JH with 5 mM 20HE suppresses mean ISC mitoses to 11.5. After 72 h of feeding, the average number of ISC mitoses per midgut was as follows. Vehicle control: 5.5, 1.5 mM methoprene: 9.5, 5 mM 20HE: 13.5 mitoses, 5 mM 20HE + 1 mM JH 10.9, 5 mM 20HE + 1.5 mM methoprene: 10. These results indicate that 16 h of 2–5 mM 20HE act as a strong promitotic signal to ISCs of virgin females, but after 72 h the mean 20HE-induced mitoses drop towards basal levels. 1.5 mM methoprene causes a mild but persistent increase in ISC mitoses over 72 h. Overnight combination feeding of 20HE and 1.5 mM methoprene or 1 mM JH strongly suppressed 20HE-induced mitoses. **e**, Methoprene does not promote ISC mitoses in mated females. Mated females were fed with 1 mM or 5 mM active JH III ligand (“JH”), JH agonist methoprene (“M”), 1 mM or 5 mM 20HE, or 20HE and JH in combination and their mitotic indexes were assessed 16-18 h after feeding. Feeding of 1 mM or 5 mM JH, 1 mM 20HE, 1.5 mM or 5 mM methoprene do not induce mitoses in mated females. 5 mM 20HE feeding induces a boost of ISC mitoses that were suppressed by combination feeding with 1 mM JH. **f**, Exogenous JH feeding inhibits ISC mitoses when combined with other promitotic stimuli. Mated females were heat-shocked for 30 min, infected with *P.e.* for 18–20 h or fed with 20HE, either alone or in combination with 1 mM JH feeding for 16–18 h, and mitotic indexes were scored. In each case, feeding 1 mM JH suppresses the mitotic response of the stimulus. **g**, Ovarian ecdysteroidogenic enzymes are required for methoprene-induced mitoses of the midgut. 1.5 mM methoprene causes ISC mitoses in control midguts (mean of 6.5 mitoses relative to 2 mitoses in vehicle-fed flies). In animals in which the ecdysteroidogenic enzyme Dib is depleted in ovaries, methoprene failed to significantly induce ISC proliferation (mean of 3.3 mitoses relative to mean of 1.4 basal mitoses in dib^{RNAi} vehicle-fed flies). Virgins were aged for 8 days at permissive temperature then fed with the different hormone regimes for 3 days. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. Results in dot plots are from three independent biological replicates. $n \geq 10$ are plotted for each genotype in the remaining scatter plots. Data are mean ± s.d. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, ****$P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the online Source Data. Representative images are shown from experiments that were repeated three independent times.

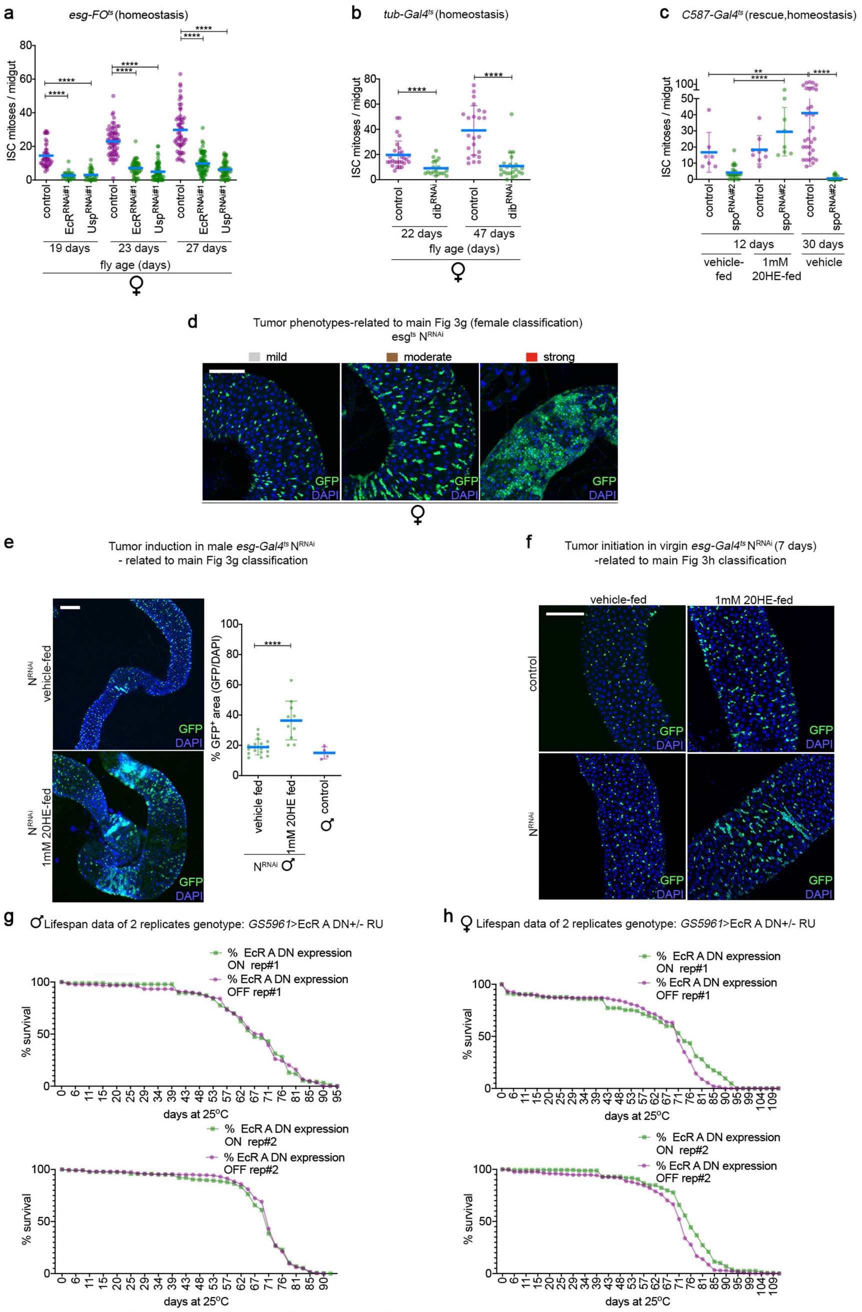


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Eip75B is a downstream ecdysone-inducible effector required to stimulate ISC proliferation, through Hr3 repression. a, 20HE feeding or *P.e.* infection transcriptionally upregulate the ecdysone-inducible targets Eip75B and Broad. 5–7-day-old mated females were fed with 20HE or infected with *P.e.* for 6 h, then mRNA levels were determined by RT–qPCR on RNA from whole midguts. Expression is indicated as mean fold change relative to vehicle-treated midguts ± s.d. ($n = 4$). **b**, Broad and Eip75B are required by adult *Drosophila* midgut progenitors for *P.e.* or 20HE-induced ISC mitoses. Increased mitoses were observed after *P.e* infection or 20HE feeding in control mated flies, which were significantly blunted after Broad or Eip75B depletion in midgut progenitors. **c**, **d**, Eip75B is only cell-autonomously required in ISCs (**c**), but not EBs (**d**) for *P.e.*-or 20HE-induced ISC mitoses. Flies were fed with 20HE or *P.e.* for 16–20 h. Results are shown for two independent RNAi lines. **e**, Eip75B-null mutant clones are strongly impaired in their ability to divide and regenerate the epithelium. Eip75B-null mutant clones were generated by MARCM and analysed 6 days after *P.e.* infection. This experiment was done with a different recombinant mutant stock than that used in Extended Data Fig. 2. **f**, *Eip75B*^{RNAi} blocks renewal of the midgut epithelium; *br*^{RNAi} does not. Representative images from ISC clones of ageing epithelia with reduced levels of Eip75B or Broad. Broad depletion does not affect ISC clonal growth, whereas Eip75B depletion blocks any ISC growth and most cells remain singlets. **g**, Eip75B overexpression in ISC-derived *esgFO*^{ts} clones is pro-proliferative as shown by representative images of ISC clones in the epithelium of mated females. **h**, **i**, Eip75B is required by ISCs to divide in response to 20HE, haem, paraquat and enteric infection. **h**, Representative images of Eip75B-depleted ISC clones in response to the different stresses. Clonal growth to any stress stimulus is impaired. **i**, Quantification of mitotic counts. Results for *P.e.*-induced mitoses are shown for two independent *Eip75B*^{RNAi} lines. **j**, Representative images of the heatshock-inducible Hr3 reporter (*hs-Gal4.DBD-Hr3.LBD>GFP*). Conditions of low Eip75B activity result in high Hr3 reporter expression and high Eip75B activity is reflected by low Hr3 reporter expression. Of note, owing to its transcriptional repressive activity, the Eip75B reporter cannot be used to monitor its activity[30]. Under basal conditions, midguts express high levels of Hr3 reporter. Hr3 activity is repressed by 20HE or haem feeding, *P.e.* infection (stimuli that require Eip75B) or co-expression of *Eip75B*. Nitric oxide (NO) inhibits Eip75B binding to Hr3[37]. SNAP is a nitric oxide donor compound that modulates nitric oxide availability and is used to regulate Eip75B activity. However, increased nitric oxide levels through SNAP feeding relieved the repressive actions of *P.e.* and *Eip75B* on GFP expression. This indicates that in ISCs, Eip75B inhibits Hr3 and nitric oxide blocks this suppressive effect. Right, mitotic counts are shown for vehicle-fed, haem-fed, *P.e.*, or *P.e.*+SNAP-fed mated females after 30 min heatshock (to induce the Hr3-GFP reporter) and 18–20 h of feeding. **k**, **l**, Hr3 overexpression strongly impairs epithelial renewal as the flies age, depicted by quantifications of mitotic indexes in **k**. **l**, Representative images of GFP-marked Hr3-overexpressing ISC clones showing impaired clonal growth in midguts of mated females. **m**, Hr3 depletion permits ISCs to divide in response to *P.e.* infection as shown by mitotic counts of Hr3-depleted ISC clones in mated females, which respond to *P.e.* infection at similar rates to control midguts. **n**, Repression of ISC mitoses in Eip75B-depleted *esgFO*^{ts} clones is rescued by *Hr3*^{RNAi} as shown by mitotic counts of ageing or *P.e.*-infected guts with Eip75B, Hr3 depletion or both. This experiment shows that Hr3 is epistatic to Eip75B. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. The overnight standard period of feeding the flies was 18–20 h. Results in dot plots are from three independent biological replicates. $n \geq 10$ are plotted for each genotype in the scatter plots. Data are mean ± s.d. $^{****}P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the online Source Data. Representative images are shown from experiments that were repeated three times. Scale bars, 100 µm.



Extended Data Fig. 9 | Nitric oxide modulates the interaction of Eip75B and Hr3 to regulate ISC division. a–**d**, Eip75B is not required in other midgut cell types besides progenitors for *P.e.* infection to induce ISC proliferation. Eip75B was depleted in progenitors using *esg-gal4^{ts}* (two independent RNAi lines are shown to complement results in Fig. 2) (**a**), visceral muscle using *how-Gal4^{ts}* (**b**), ECs using *Myo1A-gal4^{ts}* (**c**), or enteroendocrine cells using *prosV1-gal4^{ts}* (**d**). **e**, Overexpression of Hr3 in ISC-derived clones impedes the mitotic ability of ISCs to divide in response to *P.e.* infection. **f**, **g**, Inhibition of nitric oxide (NO) rescues the ISC mitotic activity of Hr3-overexpressing progenitors. **f**, ISC mitotic counts. **g**, Representative images of progenitor-specific overexpression of GFP with or without Hr3 followed by *P.e.* infection alone or in combination with the NO inhibitor L-NAME. NO represses the ability of Eip75B to interact with Hr3 hence, allowing transcriptional regulation of Hr3 targets. Treatment with L-NAME rescued the ISC ability to divide and progenitors expanded to fill the epithelium similar to the control mated females after infection (compare to results in Extended Data Fig. 8j). **h**, Model summarizing the regulation of Eip75B, Hr3 and Broad. **i**, Model summarizing the crosstalk between the gut and the ovary. For all panels, control flies express *UAS-GFP* instead of the transgene. The period of RNAi induction is indicated. The overnight standard period of feeding the flies was 18–20 h. Results in dot plots are from three independent biological replicates. $n \geq 10$ are plotted for each genotype in the scatter plots. Data are mean ± s.d. ****$P < 0.0001$, Mann–Whitney test with two-tailed distribution. Exact *n* numbers and *P* values are in the online Source Data. Representative images are shown from experiments that were repeated three times. Scale bars, 100 µm.



Condition	X^2	P-value
EcR A DN expression OFF rep#1 v.s. EcR A DN expression ON rep#1	0.01	0.9301
Condition	X^2	P-value
EcR A DN expression OFF rep#2 v.s. EcR A DN expression ON rep#2	0.240	0.625

Condition	X^2	P-value
EcR A DN expression OFF rep#1 v.s. EcR A DN expression ON rep#1	11.330	0.001
Condition	X^2	P-value
EcR A DN expression OFF rep#2 v.s. EcR A DN expression ON rep#2	17.730	0.000026

i

Collective life span analyses (% mortality)

♂					Age in days at % mortality				
Name	No. of subjects	Days	S.E.	95% C.I.	25%	50%	75%	90%	100%
EcR A DN expression OFF rep#1	119	65.60	1.58	62.51 ~ 68.69	57.00	69.00	76.00	83.00	90.00
EcR A DN expression ON rep#1	94	66.12	1.58	63.03 ~ 69.20	57.00	67.00	78.00	83.00	95.00
EcR A DN expression OFF rep#2	243	69.42	0.84	67.77 ~ 71.08	67	71	76	78	90
EcR A DN expression ON rep#2	263	67.98	0.9	66.21 ~ 69.75	67	71	76	81	92
♀					Age in days at % mortality				
Name	No. of subjects	Days	S.E.	95% C.I.	25%	50%	75%	90%	100%
EcR A DN expression OFF rep#1	161	63.39	1.86	59.75 ~ 67.04	60	71	78	81	90
EcR A DN expression ON rep#1	117	64.62	2.5	59.72 ~ 69.51	55	74	83	90	95
EcR A DN expression OFF rep#2	247	68.90	1.05	66.84 ~ 70.96	67.00	74.00	78.00	83.00	99.00
EcR A DN expression ON rep#2	173	74.26	1.12	72.07 ~ 76.46	71.00	76.00	83.00	90.00	111.00

Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Ovary-derived 20HE promotes intestinal dysplasia through EcR, Usp and Eip75B, which may affect *Drosophila* lifespan. a, The number of mitotic cells in midguts increases with age, and this is inhibited by RNAi-mediated knockdown of EcR or Usp in ISC clones (*esgFO*^{ts}). Mitotic counts are shown at 19, 23 and 27 days after eclosion in non-stressed female guts. **b**, Basal 20HE levels promote age-dependent intestinal dysplasia. Mitotic indexes are shown in aged mated female midguts from flies ubiquitously expressing *dib*^{RNAi} at two different ages after RNAi induction. **c**, Ovary ecdysone is required for ISC mitoses in non-stressed animals. Young and old mated females with *spo* knockdown in their ovaries have reduced ISC mitoses compared to controls. This was rescued by feeding the flies 1 mM 20HE. A second independent RNAi for *spo* is shown to complement data in Fig. 2. **d**, Representative images for the three classes of tumour phenotypes used to score mated female tumours in Fig. 3. **e**, 20HE feeding potentiates the tumour growth in *N*^{RNAi} males. Left, representative images with which males have been scored in Fig. 3. Males exhibiting big tumour clusters of at least 30 neighbouring cells along the gut were classified strong. By contrast, guts with one or two tumour clusters with less than ten neighbouring cells were classified mild. Right, quantifications are derived by calculating the ratio between the GFP^{+} area and DAPI^{+} area. Tumour induction was commenced a few days before 20HE feeding. **f**, 20HE feeding potentiates the tumour initiation in virgin females with *N*^{RNAi}. Representative images are shown for the quantifications presented in Fig. 3. Guts with no tumour clusters and just doublets of progenitor cells were classified as mild. Guts with tumour clusters of fewer than 10 neighbouring cells were classified as moderate, and guts with tumour clusters of at least 30 neighbouring cells were classified as strong. **g**–**i**, Progeny of the *GS5961-Gal4 UAS-EcR*^{F645A} genotype were mated for 48 h. The populations followed up were segregated based on their sex (males (**g**) and females (**h**)) and separated into groups of 25 flies per vial. Approximately half of the flies were fed 0.2 mg ml^{-1} RU486 to induce dominant-negative EcR expression in progenitors and the other half were fed with vehicle. RU486 or vehicle (ethanol) was deposited on the food vials 4–6 h before flipping the flies into the vials at 48-h intervals. Dead flies were visually identified and recorded. Lifespan assays were performed in two replicates and for each replicate the percentage survival was plotted as a function of days elapsed after the start of the assay. Statistical analysis was performed using log-rank test. χ^2 represents chi-squared value and the *P* values were provided from pairwise comparison with Bonferroni correction. **i**, Experimental details and the percentage mortality of the male or female replicates. Exact *n* numbers and *P* values are in the online Source Data.

Corresponding author(s): Bruce A. Edgar, Aurelio A. Teleman

Last updated by author(s): Mar 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Image data were collected using LAS AF software (version 2.7.3)for Leica DM and SP5 microscopes. RT-qPCR data were collected using Roche Light Cycler 480 (software version 1.5) and StepOnePlus Real-Time PCR System Appliedbiosystems (StepOne software version 2.0). Life span data was analyzed using Oasis (version 2).
Data analysis	Data were collated using Microsoft Excel (v16.33). Data was then entered into GraphPad Prism (v8.3.1), which was used to derive p values, standard deviations, to perform indicated statistical tests, and to generate the graphs displayed in all figures. Image files were analyzed in FIJI image analysis software (version 2.0.0-rc-69/1.52p), using custom macros as indicated. All custom macros are available for download as supplementary files or from the authors.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The source data for all figures is available via the online Supplementary Information.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on practical considerations (e.g. number of animals available, processing time required) and expected variations between animals for a given sample. Based on these considerations, sample sizes were general set at >20 animals per genotype/condition, which allowed modest standard deviations and highly significant p values (p<0.0001) to be obtained.
Data exclusions	No data were excluded.
Replication	All experiments were replicated three or more times. In many cases, data from each of the three replicates were pooled for display. In some cases, data from one or two representative replicates are displayed.
Randomization	Data were not randomized. Each sample was defined by a specific genotype and condition, and all animals from any specific genotype/condition were analyzed and included in the dataset.
Blinding	Many samples were scored manually by observation with a microscope. These include experiments in which PH3+ cells/gut were counted. These samples, as well as those in which tumor quantifications or gut sizes (area, width, length) were measured or scored blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies: Chicken anti-GFP (Life Technologies/Molecular probes, 1:500); rabbit anti-phospho-Histone 3 (Merck Millipore 1:1000); mouse anti-phospho-Histone 3 (Cell Signaling, 1:1000); guinea pig anti-GFP (Teleman Lab, 1:1000); Chicken anti-beta-galactosidase (Abcam, 1:1000); Rabbit phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) #9101 (Cell Signaling, 1:400). Secondary antibodies: (Alexa 488, 568 or 633, Invitrogen).
Validation	All antibodies were validated on Drosophila tissue samples known to have positive and negative signals based on genotype and condition.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Drosophila melanogaster were used for all experiments. Sexes and ages are noted in the Methods section, with sex and age details for each experiment listed in the Figure Legends and Figures. The precise genotypes and sources for each genetic strain used for each experiment are listed in Supplementary Tables S1 and S2.
Wild animals	This study did not use wild animals. All Drosophila stocks were obtained from public stock centers or other research labs, as listed in Supplementary Table S2.
Field-collected samples	This study did not use field-collected samples.

Ethics oversight	These studies were performed at the German Cancer Research Center (DKFZ; Heidelberg), which provides ethics guidance for laboratory animal studies according to German law. However, no ethical guidance was required for our work with Drosophila melanogaster, an invertebrate insect.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Reconstruction of the full transmission dynamics of COVID-19 in Wuhan

https://doi.org/10.1038/s41586-020-2554-8

Received: 14 April 2020

Accepted: 10 July 2020

Xingjie Hao[1,2,8], Shanshan Cheng[1,2,8], Degang Wu[1,2,8], Tangchun Wu[1,3,4] ✉, Xihong Lin[5,6,7] ✉ & Chaolong Wang[1,2,4] ✉

As countries in the world review interventions for containing the pandemic of coronavirus disease 2019 (COVID-19), important lessons can be drawn from the study of the full transmission dynamics of its causative agent—severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)— in Wuhan (China), where vigorous non-pharmaceutical interventions have suppressed the local outbreak of this disease[1]. Here we use a modelling approach to reconstruct the full-spectrum dynamics of COVID-19 in Wuhan between 1 January and 8 March 2020 across 5 periods defined by events and interventions, on the basis of 32,583 laboratory-confirmed cases[1]. Accounting for presymptomatic infectiousness[2], time-varying ascertainment rates, transmission rates and population movements[3], we identify two key features of the outbreak: high covertness and high transmissibility. We estimate 87% (lower bound, 53%) of the infections before 8 March 2020 were unascertained (potentially including asymptomatic and mildly symptomatic individuals); and a basic reproduction number (R_0) of 3.54 (95% credible interval 3.40–3.67) in the early outbreak, much higher than that of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS)[4,5]. We observe that multipronged interventions had considerable positive effects on controlling the outbreak, decreasing the reproduction number to 0.28 (95% credible interval 0.23–0.33) and—by projection—reducing the total infections in Wuhan by 96.0% as of 8 March 2020. We also explore the probability of resurgence following the lifting of all interventions after 14 consecutive days of no ascertained infections; we estimate this probability at 0.32 and 0.06 on the basis of models with 87% and 53% unascertained cases, respectively—highlighting the risk posed by substantial covert infections when changing control measures. These results have important implications when considering strategies of continuing surveillance and interventions to eventually contain outbreaks of COVID-19.

COVID-19, caused by SARS-CoV-2, was detected in Wuhan in December 2019[6]. The high population density, together with increased social activities before the Chinese New Year, catalysed the outbreak; the spread of the outbreak was expedited by massive human movement during the *Chunyun* holiday travel season from 10 January 2020[3]. Shortly after the confirmation of human-to-human transmission, the Chinese authorities implemented an unprecedented cordon sanitaire of Wuhan on 23 January to contain the geographical spread of the disease, followed by a series of non-pharmaceutical interventions—including suspension of all intra- and inter-city transportation, compulsory mask wearing in public places, cancellation of social gatherings and the home quarantine of individuals with presumed infections, those with COVID-19 related symptoms and their close contacts[1]—to reduce virus transmission. From 2 February, a strict stay-at-home policy for all residents, and the centralized isolation and quarantine of all patients, individuals suspected to have contracted the virus and their close contacts were implemented to stop household and community transmission. In addition, a city-wide door-to-door universal survey of symptoms was carried out during 17–19 February by designated community workers, to identify previously undetected symptomatic cases. These interventions—together with improved medical resources and the redeployment of healthcare personnel from all over the country—have crushed the epidemic curve and reduced the attack rate in Wuhan, with the potential to shed light on global efforts to control outbreaks of COVID-19[1].

Recent studies have revealed important transmission features of COVID-19, including the infectiousness of asymptomatic[7–10] and

---

[1] Ministry of Education Key Laboratory of Environment and Health, State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [2] Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [3] Department of Occupational and Environmental Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [4] National Medical Center for Major Public Health Events, Huazhong University of Science and Technology, Wuhan, China. [5] Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [6] Department of Statistics, Harvard University, Cambridge, MA, USA. [7] Broad Institute of MIT and Harvard, Cambridge, MA, USA. [8] These authors contributed equally: Xingjie Hao, Shanshan Cheng, Degang Wu. ✉e-mail: wut@mails.tjmu.edu.cn; xlin@hsph.harvard.edu; chaolong@hust.edu.cn

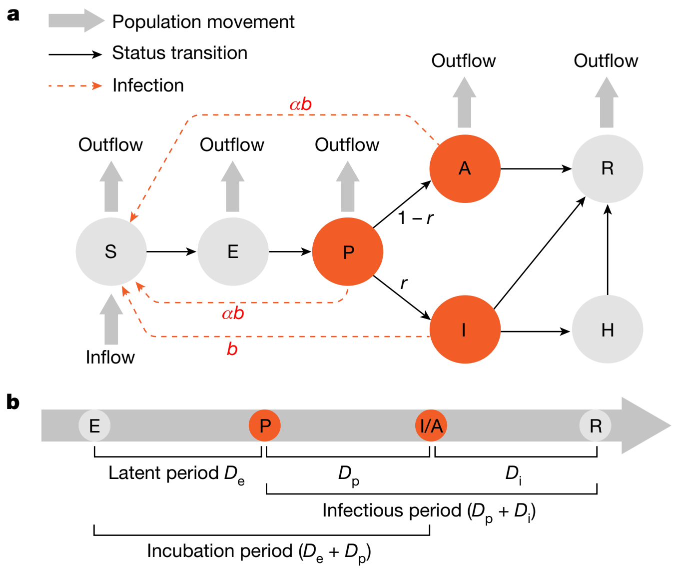


Fig. 1 | Illustration of the SAPHIRE model. We extended the classic SEIR model to include seven compartments: susceptible (S), exposed (E), presymptomatic infectious (P), ascertained infectious (I), unascertained infectious (A), isolation in hospital (H) and removed (R). **a**, Relationship between different compartments. Two parameters of interest are r (ascertainment rate) and b (transmission rate), which are assumed to vary across time periods. **b**, Schematic disease course of symptomatic individuals. In this model, the unascertained compartment A includes asymptomatic and some mildly symptomatic individuals who were not detected. Although there is no presymptomatic phase for asymptomatic individuals, we treated asymptomatic as a special case of mildly symptomatic and modelled both with a 'presymptomatic' phase for simplicity.

presymptomatic[2,11,12] individuals. Furthermore, the number of ascertained cases was much smaller than that estimated using international cases exported from Wuhan before the travel suspension[3,13,14], which implies a substantial number of unascertained cases. Using reported cases from 375 cities in China, a previous modelling study concluded that a sizeable number of unascertained cases—despite having lower transmissibility—had facilitated the rapid spreading of COVID-19[15]. In addition, accounting for unascertained cases has refined the estimation of case fatality risk of COVID-19[16]. Modelling both ascertained and unascertained cases is important for interpreting transmission dynamics and epidemic trajectories.

On the basis of comprehensive epidemiological data from Wuhan[1], we delineated the full dynamics of COVID-19 in the epicentre by extending the susceptible–exposed–infectious–recovered (SEIR) model to include presymptomatic infectiousness (P), unascertained cases (A) and case isolation in the hospital (H), generating a model that we name SAPHIRE (Fig. 1, Methods, Extended Data Tables 1, 2). We modelled the outbreak from 1 January 2020 across 5 time periods that were defined on the basis of key events and interventions: 1–9 January (before *Chunyun*), 10–22 January (*Chunyun*), 23 January to 1 February (cordon sanitaire), 2–16 February (centralized isolation and quarantine) and 17 February to 8 March (community screening). We assumed a constant population size of 10 million with equal numbers of daily inbound and outbound travellers (500,000 before *Chunyun*, 800,000 during *Chunyun* and 0 after cordon sanitaire)[3]. Furthermore, we assumed that the transmission rate and ascertainment rate did not change in the first two periods (because few interventions were implemented before 23 January), whereas these rates were allowed to vary in later periods to reflect the strengths of different interventions. We estimated these rates across periods by Markov Chain Monte Carlo (MCMC) and further converted the transmission rate into the effective reproduction number (R_e) (Methods).

We first simulated epidemic curves with two periods to validate our parameter estimation procedure (Methods, Extended Data Fig. 1). Our method could accurately estimate R_e and the ascertainment rates when the model was correctly specified, and was robust to misspecification of the duration from the onset of symptoms to isolation and of the relative transmissibility of unascertained versus ascertained cases. As expected, estimates of R_e were positively correlated with the specified latent and infectious periods, and the estimated ascertainment rates were positively correlated with the specified ascertainment rate in the initial state.

Using confirmed cases exported from Wuhan to Singapore (Extended Data Table 3), we conservatively estimated the ascertainment rate during the early outbreak in Wuhan to be 0.23 (95% confidence interval 0.14–0.42; unless specified otherwise, all parenthetical ranges refer to the 95% credible interval) (Methods). We then fit the daily incidences in Wuhan from 1 January to 29 February, assuming the initial ascertainment rate was 0.23, and predicted the trend from 1 March to 8 March (Methods). Our model fit the observed data well, except for the outlier on 1 February; this outlier might be due to the approximate-date records of many patients admitted to the field hospitals set up after 1 February (Fig. 2a). After a series of multifaceted public health interventions, R_e decreased from 3.54 (3.40–3.67) and 3.32 (3.19–3.44) in the first two periods to 1.18 (1.11–1.25), 0.51 (0.47–0.54) and 0.28 (0.23–0.33) in the later three periods (Fig. 2b, Extended Data Tables 4, 5). We estimated the cumulative number of infections, including unascertained cases, up until 8 March to be 258,728 (204,783–320,145) if the trend of the fourth period was assumed (Fig. 2c), 818,724 (599,111–1,096,850) if the trend of the third period was assumed (Fig. 2d) or 6,302,694 (6,275,508–6,327,520) if the trend of the second period was assumed (Fig. 2e), in comparison to the estimated total infections of 249,187 (198,412–307,062) obtained by fitting data from all 5 periods (Fig. 2a). Correspondingly, these numbers translate into a 3.7%, 69.6% and 96.0% reduction of infections by the measures taken in the fifth period, the fourth and the fifth periods combined, and the last three periods combined, respectively.

We estimated low ascertainment rates throughout: 0.15 (0.13–0.17) for the first two periods, and 0.14 (0.11–0.17), 0.10 (0.08–0.12), and 0.16 (0.13–0.21) for the remaining three periods (Extended Data Table 6). Even with the universal screening of the community for symptoms that was implemented from 17 February to 19 February, the ascertainment rate was raised only to 0.16. On the basis of the fitted model using data from 1 January to 29 February, we projected the cumulative number of ascertained cases to be 32,577 (30,216–34,986) by 8 March, close to the reported number of 32,583. This was equivalent to an overall ascertainment rate of 0.13 (0.11–0.16) given the estimated total infections of 249,187 (198,412–307,062). The model also projected that the number of daily active infections (including presymptomatic, ascertained and unascertained infections) peaked at 55,879 (43,582–69,571) on 2 February and dropped afterwards to 701 (436–1,043) on 8 March (Fig. 2f). If the trend remained unchanged, the number of ascertained infections would have first become zero on 27 March (95% credible interval 20 March to 5 April), and the clearance of all infections would have occurred on 21 April (8 April to 12 May) (Extended Data Table 7). The first day of zero ascertained cases in Wuhan was reported on 18 March, indicating enhanced interventions in March.

We used stochastic simulations to investigate the implications of unascertained cases for continuing surveillance and interventions[17] (Methods). Because of latent, presymptomatic and unascertained cases, the source of infection would not be completely cleared shortly after the first day of zero ascertained cases. We found that if control measures were lifted 14 days after the first day of zero ascertained cases, the probability of resurgence, defined as the number of active ascertained cases greater than 100, could be as high as 0.97, and the surge was predicted to occur on day 34 (27–47) after lifting controls (Fig. 3). If we were to impose a more-stringent criterion of lifting controls after

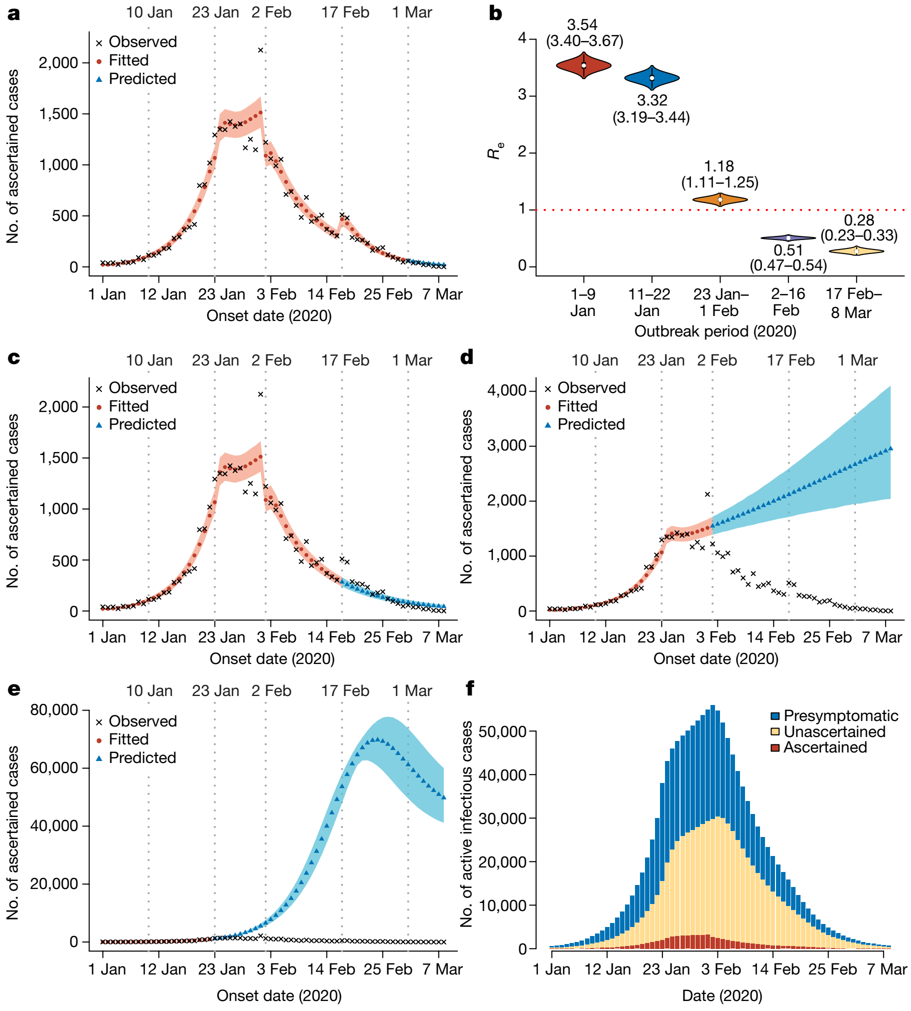


Fig. 2 | Modelling the COVID-19 epidemic in Wuhan. Parameters were estimated by fitting data from 1 January to 29 February. **a**, Prediction using parameters from period 5 (17 February–29 February). **b**, Distribution of R_e estimates from 10,000 MCMC samples. In each violin plot, the white dot represents the median, the thick bar represents the interquartile range and the thin bar represents the minimum and the maximum. The mean and the 95% credible interval (in parentheses) are labelled below or above. **c**, Prediction using parameters from period 4 (2 February–16 February). **d**, Prediction using parameters from period 3 (23 January to 1 February). **e**, Prediction using parameters from period 2 (10 January to 22 January). The shaded areas in **a**, **c**–**e** are 95% credible intervals, and the coloured points are the mean values based on 10,000 MCMC samples. **f**, Estimated number of active infectious cases in Wuhan from 1 January to 8 March.

observing no ascertained cases in a consecutive period of 14 days, the probability of resurgence would drop to 0.32, with possible resurgence delayed to day 42 (33–55) after lifting controls (Fig. 3). These results highlight the risk of ignoring unascertained cases in switching intervention strategies, despite our use of a simplified model.

We performed a series of sensitivity analyses to test the robustness of our results by smoothing the outlier data point on 1 February, as well as varying the lengths of latent and infectious periods, the duration from the onset of symptoms to isolation, the ratio of transmissibility in unascertained versus ascertained cases, and the initial ascertainment rate (Extended Data Tables 4–7, Supplementary Information). Our major findings, of a marked decrease in R_e after interventions and the existence of a substantial number of unascertained cases, were robust. Consistent with simulations, the estimated ascertainment rates were positively correlated with the specified initial ascertainment rate. When we specified the initial ascertainment rate as 0.14 or 0.42, the estimated overall ascertainment rate was 0.08 (0.07–0.10) and 0.23 (0.16–0.28), respectively. If we assume an extreme scenario with no unascertained cases in the early outbreak (which we term model 'S8' (Supplementary Information)), the estimated ascertainment rate would be 0.47 (0.39–0.58) overall, which would represent an upper bound of the ascertainment rate. Because of the higher ascertainment rate (compared to the main analysis) in this model, we estimated a lower probability of resurgence (0.06) when lifting controls after 14 days of no ascertained cases, and the resurgence was expected to occur on day 38 (29–52) after lifting controls (Fig. 3). A simplified model that assumes complete ascertainment at any time performed substantially worse than the full model (Extended Data Table 4, Supplementary Information).

Understanding the proportion of unascertained cases and their transmissibility is critical for the prioritization of the surveillance and control measures[17]. Our finding of a large fraction of unascertained cases—despite the high level of surveillance in Wuhan—indicates the existence of many asymptomatic or mildly symptomatic individuals. It was previously estimated that asymptomatic individuals accounted for 18% of the infections on board the Diamond Princess Cruise ship[8] and 31% of the infected Japanese individuals who were evacuated from Wuhan[9]. In addition, in a cohort of 210 women admitted for delivery between 22 March and 4 April in New York City (USA), 29 of 33 (88%) pregnant women infected with SARS-CoV-2 were asymptomatic[10]. Several reports have also highlighted the difficulty of detecting cases of COVID-19: the detection capacity varied from 11% in low-surveillance

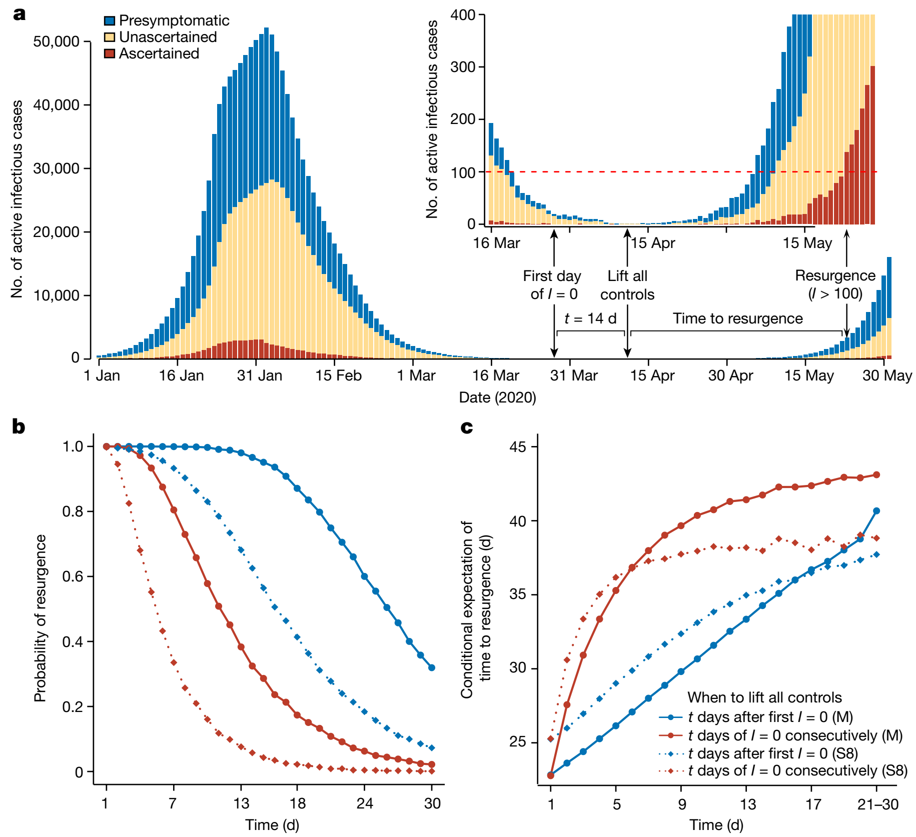


Fig. 3 | Risk of resurgence after lifting controls. We consider the main model (M) and the sensitivity analysis (S8) (Methods). In model M, we assume the initial ascertainment rate $r_0 = 0.23$, and thus an overall ascertainment rate of 0.13. In S8, we assume no unascertained cases initially and thus an overall ascertainment rate of 0.47. For each model, we simulated epidemic curves on the basis of 10,000 sets of parameters from MCMC, and set the transmission rate (*b*), ascertainment rate (*r*) and population movement (*n*) to their values in the first period after lifting controls. Resurgence was defined as reaching over 100 active ascertained infections. **a**, Illustration of a simulated curve under the main model, with control measures lifted 14 days after the first day of no ascertained cases. The inset is an enlarged plot from 16 March to 28 May. **b**, Probability of resurgence if control measures were lifted *t* days after the first day of no ascertained cases, or after observing zero ascertained cases for *t* days consecutively. **c**, Expectation of time to resurgence, conditional on the occurrence of resurgence. We grouped the final 10 days ($t = 21$ to 30) to calculate the expected time to resurgence because of their low probability of resurgence. Key applies to both **b** and **c**.

countries to 40% in high-surveillance countries[18,19], and the modelling of epidemics outside of Wuhan has suggested that the ascertainment rate was 24.4% in China (excluding Hubei province)[14] and 14% in Wuhan before the travel ban[15]. Consistent with these studies and emerging serological studies that show that seroprevalence is much higher than the reported case prevalence in cities and countries worldwide[20–22], our analyses of data from Wuhan indicated an overall ascertainment rate between 8% and 23% (Extended Data Table 6, excluding the extreme scenario of model S8).

Our R_e estimate of 3.54 (3.40–3.67) before any interventions is at the higher end of the range of the estimated R_0 values of other studies that used early epidemic data from Wuhan[6,23]. This discrepancy might be due to the modelling of unascertained cases, more-complete case records in our analysis and/or to the different time periods analysed. If we modelled from the first case of COVID-19 reported in Wuhan, we would estimate a lower R_e of 3.38 (3.28–3.48) before interventions (Extended Data Fig. 2), which remains much higher than those of SARS and MERS[4,5].

Our modelling study has delineated the full-spectrum dynamics of the COVID-19 outbreak in Wuhan, and highlighted two key features of the outbreak: high covertness and high transmissibility. These two features have synergistically propelled the COVID-19 pandemic, and imposed considerable challenges to attempts to control the outbreak. However, the Wuhan case study demonstrates the effectiveness of vigorous and multifaceted containment efforts. In particular, despite the relatively low ascertainment rates (owing to mild or absent symptoms of many infected individuals), the outbreak was controlled by interventions such as wearing face masks, social distancing and quarantining close contacts[1], which block transmission that stems from unascertained cases.

Given the limitations of our model as discussed below, further investigations—such as a survey of the seroprevalence of SARS-CoV-2-specific antibodies—are needed to confirm our estimates. First, owing to the delay in laboratory tests, we might have missed some cases and therefore underestimated the ascertainment rate (especially for the last period). Second, we excluded clinically diagnosed cases without laboratory confirmation to reduce false-positive diagnoses; however, this leads to an underestimation of ascertainment rates—especially for the third and fourth periods, during which many clinically diagnosed cases were reported[1]. The variation in the estimated ascertainment rates across periods reflects a combined effect of the evolving surveillance, interventions, medical resources and case definitions across time periods[1,24]. Third, our model assumes homogeneous transmission within the population and ignores heterogeneity between groups by sex, age, geographical region and socioeconomic status[25]. Furthermore, individual variation in infectiousness—such as superspreading events[26]—is known to result in a higher probability of stochastic extinction given a fixed population R_e (ref. [27]). We might therefore have overestimated the probability of resurgence. Finally, we could not evaluate the effect of individual interventions on the basis of an epidemic curve from a single city, because many interventions were applied simultaneously. Future work that models heterogeneous transmission between different groups, and joint analysis with data from other cities, will provide deeper insights into the effectiveness of different control strategies[28,29].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2554-8.

1. Pan, A. et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *J. Am. Med. Assoc.* **323**, 1915–1923 (2020).
2. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
3. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
4. Wang, Y., Wang, Y., Chen, Y. & Qin, Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J. Med. Virol.* **92**, 568–576 (2020).
5. Lipsitch, M. et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970 (2003).
6. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
7. Bai, Y. et al. Presumed asymptomatic carrier transmission of COVID-19. *J. Am. Med. Assoc.* **323**, 1406–1407 (2020).
8. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25**, 2000180 (2020).
9. Nishiura, H. et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155 (2020).
10. Sutton, D., Fuchs, K., D'Alton, M. & Goffman, D. Universal screening for SARS-CoV-2 in women admitted for delivery. *N. Engl. J. Med.* **382**, 2163–2164 (2020).
11. Tong, Z. D. et al. Potential presymptomatic transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerg. Infect. Dis.* **26**, 1052–1054 (2020).
12. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
13. Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558 (2020).
14. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
15. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
16. Wu, J. T. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
17. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the epidemiology of COVID-19 – studies needed. *N. Engl. J. Med.* **382**, 1194–1196 (2020).
18. De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. O. & Lipsitch, M. Identifying locations with possible undetected imported severe acute respiratory syndrome coronavirus 2 cases by using importation predictions. *Emerg. Infect. Dis.* **26**, 1465–1469 (2020).
19. Niehus, R., De Salazar, P. M., Taylor, A. R. & Lipsitch, M. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *Lancet Infect. Dis.* **20**, 803–808 (2020).
20. Levesque, J. & Maybury, D. W. A note on COVID-19 seroprevalence studies: a meta-analysis using hierarchical modelling. Preprint at https://doi.org/10.1101/2020.05.03.20089201 (2020).
21. To, K. K.-W. et al. Seroprevalence of SARS-CoV-2 in Hong Kong and in residents evacuated from Hubei province, China: a multicohort study. *Lancet Microbe* **1**, E111–E118 (2020).
22. Xu, X. et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat. Med.* https://doi.org/10.1038/s41591-020-0949-6 (2020).
23. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021 (2020).
24. Tsang, T. K. et al. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* **5**, e289–e296 (2020).
25. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
26. Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **395**, e47 (2020).
27. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
28. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020).
29. Prem, K. et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).

Methods

Data of cases of COVID-19 in Wuhan

We analysed the daily incidence data of COVID-19, presented in figure 1 of ref. [1]. In brief, information on cases of COVID-19 from 8 December 2019 to 8 March 2020 were extracted from the municipal Notifiable Disease Report System on 9 March 2020. The date of the onset of symptoms (the self-reported date of developing symptoms, such as a fever, cough or other respiratory symptoms) and the date of confirmed diagnosis were collected. For the consistency of case definition throughout the periods, we included only 32,583 individuals who had a laboratory-confirmed positive test for SARS-CoV-2 by the real-time reverse-transcription polymerase-chain-reaction (RT–PCR) assay or high-throughput sequencing of nasal and pharyngeal swab specimens. SAS software (version 9.4) was used in data collection.

Estimation of initial ascertainment rate using cases exported to Singapore

As of 10 May 2020, a total of 24 confirmed cases of COVID-19 in Singapore were reported to be imported from China, among which 16 were imported from Wuhan before the cordon sanitaire on 23 January; the first case arrived in Singapore on 18 January (Extended Data Table 3). Based on VariFlight Data (https://data.variflight.com/en/), the total number of passengers who travelled from Wuhan to Singapore between 18 January and 23 January 2020 was 2,722. Therefore, the infection rate among these passengers was 0.59% (95% confidence interval 0.30–0.88%). These individuals had an onset of symptoms between 21 January and 30 January 2020. In Wuhan, a total of 12,433 confirmed cases involved individuals who were reported to have experienced an onset of symptoms in the same period—equivalent to a cumulative infection rate of 0.124% (95% confidence interval 0.122–0.126%), assuming a population size of 10 million for Wuhan. By further assuming complete ascertainment of early cases in Singapore (which is well-known for its high level of surveillance[18,19]), the ascertainment rate during the early outbreak in Wuhan was estimated to be 0.23 (95% confidence interval 0.14–0.42), corresponding to 0.77 (95% confidence interval 0.58–0.86) of the infections being unascertained. This represents a conservative estimate for two reasons: (1) the assumption of perfect ascertainment in Singapore ignored potential asymptomatic individuals;[8,9] and (2) the number of imported cases in which individuals experienced symptom onset between 21 January and 30 January was underestimated owing to the suspension of flights after lockdown in Wuhan. Without direct information to estimate the initial ascertainment rate before 1 January 2020, we used these results based on Singapore data to set the initial value and the prior distribution of ascertainment rates in our model, and performed sensitivity analyses under various assumptions.

The SAPHIRE model

We extended the classic SEIR model to a SAPHIRE model (Fig. 1, Extended Data Table 1), which incorporates three additional compartments to account for presymptomatic infectious individuals (P), unascertained cases (A) and cases isolated in the hospital (H). We chose to analyse data from 1 January 2020, when the Huanan Seafood Market was disinfected, and thus did not model the zoonotic force of infection[3]. We assumed a constant population size (N) = 10,000,000, with equal numbers of daily inbound and outbound travellers (n), in which n = 500,000 for 1–9 January, 800,000 for 10–22 January (owing to *Chunyun*) and 0 after the cordon sanitaire from 23 January[3]. We divided the population into susceptible (S), exposed (E), P, A, ascertained infectious (I), H and removed (R) individuals. We introduced compartment H because ascertained cases would have a shorter effective infectious period owing to isolation, especially when medical resources were improved[1]. We use italicized letters to denote the number of individuals in each corresponding compartment. The dynamics of these compartments across time (t) are described by the following set of ordinary differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = n - \frac{bS(\alpha P + \alpha A + I)}{N} - \frac{nS}{N} \tag{1}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{bS(\alpha P + \alpha A + I)}{N} - \frac{E}{D_\mathrm{e}} - \frac{nE}{N} \tag{2}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{E}{D_\mathrm{e}} - \frac{P}{D_\mathrm{p}} - \frac{nP}{N} \tag{3}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{(1-r)P}{D_\mathrm{p}} - \frac{A}{D_\mathrm{i}} - \frac{nA}{N} \tag{4}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{rP}{D_\mathrm{p}} - \frac{I}{D_\mathrm{i}} - \frac{I}{D_\mathrm{q}} \tag{5}$$

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{I}{D_\mathrm{q}} - \frac{H}{D_\mathrm{h}} \tag{6}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{A+I}{D_\mathrm{i}} + \frac{H}{D_\mathrm{h}} - \frac{nR}{N} \tag{7}$$

in which b is the transmission rate for ascertained cases (defined as the number of individuals that an ascertained case can infect per day); α is the ratio of the transmission rate of unascertained cases to that of ascertained cases; r is ascertainment rate; D_e is the latent period; D_p is the presymptomatic infectious period; D_i is the symptomatic infectious period; D_q is the duration from illness onset to isolation; and D_h is the isolation period in hospital. R_e could be computed as

$$R_\mathrm{e} = \alpha b\left(D_\mathrm{p}^{-1} + \frac{n}{N}\right)^{-1} + (1-r)\alpha b\left(D_\mathrm{i}^{-1} + \frac{n}{N}\right)^{-1} + rb\left(D_\mathrm{i}^{-1} + D_\mathrm{q}^{-1}\right)^{-1} \tag{8}$$

in which the three terms represent infections contributed by presymptomatic individuals, unascertained cases and ascertained cases, respectively. We adjusted the infectious periods of each type of case by taking population movement $\left(\frac{n}{N}\right)$ and isolation (D_q^{-1}) into account.

Parameter settings and initial states

Parameter settings for the main analysis are summarized in Extended Data Table 2. We set $\alpha = 0.55$ according to ref. [15], assuming lower transmissibility for unascertained cases. Compartment P contains both ascertained and unascertained cases in the presymptomatic phase. We set the transmissibility of P to be the same as unascertained cases, because it has previously been reported that the majority of cases are unascertained[15]. We assumed an incubation period of 5.2 days and a presymptomatic infectious period of $D_\mathrm{p} = 2.3$ days[2,6]. Thus, the latent period was $D_\mathrm{e} = 5.2 - 2.3 = 2.9$ days. Because presymptomatic infectiousness was estimated to account for 44% of the total infections from ascertained cases[2], we set the mean of total infectious period as $(D_\mathrm{p} + D_\mathrm{i}) = \frac{D_\mathrm{p}}{0.44} = 5.2$ days, assuming constant infectiousness across the presymptomatic and symptomatic phases of ascertained cases[12]—thus, the mean symptomatic infectious period was $D_\mathrm{i} = 2.9$ days. We set a long isolation period of $D_\mathrm{h} = 30$ days, but this parameter has no effect on our fitting procedure and the final parameter estimates. The duration from the onset of symptoms to isolation was estimated to be $D_\mathrm{q} = 21$, 15, 10, 6 and 3 days as the median time length from onset to confirmed diagnosis in period 1–5, respectively[1].

On the basis of the settings above, we specified the initial state of the model on 31 December 2019 (Extended Data Table 1). The initial number of ascertained symptomatic cases $I(0)$ was specified as the number of ascertained cases in which individuals experienced symptom onset during 29–31 December 2019. We assumed the initial ascertainment rate was r_0, and thus the initial number of unascertained cases was $A(0) = r_0^{-1}(1 - r_0)I(0)$. We denoted $P_I(0)$ and $E_I(0)$ as the numbers of ascertained cases in which individuals experienced symptom onset during 1–2 January 2020 and 3–5 January 2020, respectively. Then, the initial numbers of exposed and presymptomatic individuals were set as $E(0) = r_0^{-1}E_I(0)$ and $P(0) = r_0^{-1}P_I(0)$, respectively. We assumed $r_0 = 0.23$ in our main analysis, on the basis of the point estimate using the Singapore data (described in 'Estimation of initial ascertainment rate using cases exported to Singapore').

Estimation of parameters in the SAPHIRE model

Considering the time-varying strength of control measures, we assumed $b = b_{12}$ and $r = r_{12}$ for the first two periods, $b = b_3$ and $r = r_3$ for period 3, $b = b_4$ and $r = r_4$ for period 4, and $b = b_5$ and $r = r_5$ for period 5. We assumed that the observed number of ascertained cases in which individuals experienced symptom onset on day d—denoted as x_d—follows a Poisson distribution with rate $\lambda_d = rP_{d-1}D_p^{-1}$, in which P_{d-1} is the expected number of presymptomatic individuals on day $(d - 1)$. We fit the observed data from 1 January to 29 February ($d = 1, 2, ..., D$, and $D = 60$) and used the fitted model to predict the trend from 1 March to 8 March. Thus, the likelihood function is

$$L(b_{12}, b_3, b_4, b_5, r_{12}, r_3, r_4, r_5) = \prod_{d=1}^{D} \frac{e^{-\lambda_d}\lambda_d^{x_d}}{x_d!} \tag{9}$$

We estimated b_{12}, b_3, b_4, b_5, r_{12}, r_3, r_4 and r_5 by MCMC with the delayed rejection adaptive metropolis algorithm implemented in the R package BayesianTools (version 0.1.7)[30]. We used a non-informative flat prior of Unif(0,2) for b_{12}, b_3, b_4 and b_5. For r_{12}, we used an informative prior of Beta(7.3,24.6) by matching the first two moments of the estimate using Singapore data (described in 'Estimation of initial ascertainment rate using cases exported to Singapore'). We reparameterized r_3, r_4 and r_5 by

$$\mathrm{logit}(r_3) = \mathrm{logit}(r_{12}) + \delta_3$$

$$\mathrm{logit}(r_4) = \mathrm{logit}(r_3) + \delta_4$$

$$\mathrm{logit}(r_5) = \mathrm{logit}(r_4) + \delta_5$$

in which $\mathrm{logit}(r) = \log\left(\frac{r}{1-r}\right)$. In the MCMC, we sampled δ_3, δ_4 and δ_5 from the prior of $N(0,1)$. We set a burn-in period of 40,000 iterations and continued to run 100,000 iterations with a sampling step size of 10 iterations. We repeated MCMC with three different sets of initial values and assessed the convergence by the trace plot and the multivariate Gelman–Rubin diagnostic[31] (Supplementary Information). Estimates of parameters were presented as posterior means and 95% credible intervals from 10,000 MCMC samples. All of the analyses were performed in R (version 3.6.2) and the Gelman–Rubin diagnostic was calculated using the gelman.diag function in the R package coda (version 0.19.3).

Stochastic simulations

We used stochastic simulations to obtain the 95% credible interval of a fitted or predicted epidemic curve. Given a set of parameter values from MCMC, we performed the following multinomial random sampling:

$$(U_{S\to E}, U_{S\to O}, U_{S\to S}) \sim \mathrm{Multinomial}(S_{t-1}; p_{S\to E}, p_O, 1 - p_{S\to E} - p_O)$$

$$(U_{E\to P}, U_{E\to O}, U_{E\to E}) \sim \mathrm{Multinomial}(E_{t-1}; p_{E\to P}, p_O, 1 - p_{E\to P} - p_O)$$

$$\begin{aligned}&(U_{P\to I}, U_{P\to A}, U_{P\to O}, U_{P\to P})\\ &\quad \sim \mathrm{Multinomial}(P_{t-1}; p_{P\to I}, p_{P\to A}, p_O, 1 - p_{P\to I} - p_{P\to A} - p_O)\end{aligned}$$

$$(U_{I\to H}, U_{I\to R}, U_{I\to I}) \sim \mathrm{Multinomial}(I_{t-1}; p_{I\to H}, p_{I\to R}, 1 - p_{I\to H} - p_{I\to R})$$

$$(U_{A\to R}, U_{A\to O}, U_{A\to A}) \sim \mathrm{Multinomial}(A_{t-1}; p_{A\to R}, p_O, 1 - p_{A\to R} - p_O)$$

$$(U_{H\to R}, U_{H\to H}) \sim \mathrm{Multinomial}(H_{t-1}; p_{H\to R}, 1 - p_{H\to R})$$

$$(U_{R\to O}, U_{R\to R}) \sim \mathrm{Multinomial}(R_{t-1}; p_O, 1 - p_O)$$

in which O denotes the status of outflow population, $p_O = nN^{-1}$ denotes the outflow probability and other quantities are status transition probabilities, including $p_{S\to E} = b(\alpha P_{t-1} + \alpha A_{t-1} + I_{t-1})N^{-1}$, $p_{E\to P} = D_e^{-1}$, $p_{P\to I} = rD_p^{-1}$, $p_{P\to A} = (1 - r)D_p^{-1}$, $p_{I\to H} = D_q^{-1}$, $p_{I\to R} = p_{A\to R} = D_i^{-1}$ and $p_{H\to R} = D_h^{-1}$. The SAPHIRE model described by equations (1)–(7) is equivalent to the following stochastic dynamics:

$$S_t - S_{t-1} = n - U_{S\to E} - U_{S\to O} \tag{10}$$

$$E_t - E_{t-1} = U_{S\to E} - U_{E\to P} - U_{E\to O} \tag{11}$$

$$P_t - P_{t-1} = U_{E\to P} - U_{P\to A} - U_{P\to I} - U_{P\to O} \tag{12}$$

$$A_t - A_{t-1} = U_{P\to A} - U_{A\to R} - U_{A\to O} \tag{13}$$

$$I_t - I_{t-1} = U_{P\to I} - U_{I\to H} - U_{I\to R} \tag{14}$$

$$H_t - H_{t-1} = U_{I\to H} - U_{H\to R} \tag{15}$$

$$R_t - R_{t-1} = U_{A\to R} + U_{I\to R} + U_{H\to R} - U_{R\to O} \tag{16}$$

We repeated the stochastic simulations for all 10,000 sets of parameter values sampled by MCMC to construct the 95% credible interval of the epidemic curve by the 2.5 and 97.5 percentiles at each time point.

Prediction of epidemic ending date and the risk of resurgence

Using the stochastic simulations described in 'Stochastic simulations', we predicted the first day of no new ascertained cases and the date of clearance of all active infections in Wuhan, assuming continuation of the same control measures as the last period (that is, same parameter values).

We also evaluated the risk of outbreak resurgence after lifting control measures. We considered lifting all controls (1) at t days after the first day of zero ascertained cases, or (2) after a consecutive period of t days with no ascertained cases. After lifting controls, we set the transmission rate b, ascertainment rate r and population movement n to be the same as the first period, and continued the stochastic simulation to the stationary state. Time to resurgence was defined as the number of days from lifting controls to when the number of active ascertained cases (I) reached 100. We performed 10,000 simulations with 10,000 sets of parameter values sampled from MCMC (as described in 'Estimation of parameters in the SAPHIRE model'). We calculated the probability of resurgence as the proportion of simulations in which resurgence occurred, as well as the time to resurgence conditional on the occurrence of resurgence.

Simulation study for method validation

To validate the method, we performed two-period stochastic simulations (equations (10) to (16)) with transmission rate $b = b_1 = 1.27$, ascertainment rate $r = r_1 = 0.2$, daily population movement $n = 500{,}000$, and duration from illness onset to isolation $D_q = 20$ days for the first period (so that $R_e = 3.5$ according to equation (8)), and $b = b_2 = 0.41$, $r = r_2 = 0.4$, $n = 0$ and $D_q = 5$ for the second period (so that $R_e = 1.2$ according to equation (8)). Lengths of both periods were set to 15 days, and the initial ascertainment rate was set to $r_0 = 0.3$, and the other parameters and initial states were set as those in our main analysis (Extended Data Tables 1, 2). We repeated stochastic simulations 100 times to generate 100 datasets. For each dataset, we applied our MCMC method to estimate b_1, b_2, r_1 and r_2, and set all other parameters and initial values to be the same as the true values. We translated b_1 and b_2 into $(R_e)_1$ and $(R_e)_2$ according to equation (8), and focused on evaluating the estimates of $(R_e)_1$, $(R_e)_2$, r_1 and r_2. We also tested the robustness to misspecification of the latent period D_e, presymptomatic infectious period D_p, symptomatic infectious period D_i, duration from illness onset to isolation D_q, ratio of transmissibility between unascertained and ascertained cases α, and initial ascertainment rate r_0. In each test, we changed the specified value of a parameter (or initial state) to be 20% lower or higher than its true value, and kept all other parameters unchanged. When we changed the value of r_0, we adjusted the initial states $A(0)$, $P(0)$ and $E(0)$ according to Extended Data Table 1.

For each simulated dataset, we ran the MCMC method with 20,000 burn-in iterations and an additional 30,000 iterations. We sampled parameter values from every 10 iterations, resulting in 3,000 MCMC samples. We took the mean across 3,000 MCMC samples as the final estimates and display results for 100 repeated simulations.

Sensitivity analyses for the real data

We designed nine sensitivity analyses to test the robustness of our results from real data. For each of the sensitivity analyses, we fixed parameters and initial states to be the same as the main analysis except for those mentioned below. For analysis (S1), we adjust the reported incidences from 29 January to 1 February to their average. We suspect the spike of incidences on 1 February might be caused by approximate-date records among some patients admitted to the field hospitals after 2 February. The actual dates for illness onset for these patients were likely to be spread between 29 January and 1 February. For analysis (S2), we assume an incubation period of 4.1 days (lower 95% confidence interval from ref. [6]) and a presymptomatic infectious period of 1.1 days (the lower 95% confidence interval from ref. [2] is 0.8 days, but our discrete stochastic model requires $D_p > 1$), equivalent to set $D_e = 3$ and $D_p = 1.1$, and adjust $P(0)$ and $E(0)$ accordingly. For analysis (S3), we assume an incubation period of 7 days (upper 95% confidence interval from ref. [6]) and a presymptomatic infectious period of 3 days (upper 95% confidence interval from ref. [2]), equivalent to set $D_e = 4$ and $D_p = 3$, and adjust $P(0)$ and $E(0)$ accordingly. For analysis (S4), we assume the transmissibility of the unascertained cases is $\alpha = 0.46$ (lower 95% confidence interval from ref. [15]) of the ascertained cases. For analysis (S5), we assume the transmissibility of the unascertained cases is $\alpha = 0.62$ (upper 95% confidence interval from ref. [15]) of the ascertained cases. For analysis (S6), we assume the initial ascertainment rate is $r_0 = 0.14$ (lower 95% confidence interval of the estimate using Singapore data) and adjust $A(0)$, $P(0)$ and $E(0)$ accordingly. For analysis (S7), we assume the initial ascertainment rate is $r_0 = 0.42$ (upper 95% confidence interval of the estimate using Singapore data) and adjust $A(0)$, $P(0)$ and $E(0)$ accordingly. For analysis (S8), we assume the initial ascertainment rate is $r_0 = 1$ (theoretical upper limit) and adjust $A(0)$, $P(0)$ and $E(0)$ accordingly. For analysis (S9), we assume no unascertained cases by fixing $r_0 = r_{12} = r_3 = r_4 = r_5 = 1$. We compared this simplified model to the full model using the Bayes factor.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data analysed in this study are available on GitHub at https://github.com/chaolongwang/SAPHIRE.

Code availability

Codes are available on GitHub at https://github.com/chaolongwang/SAPHIRE.

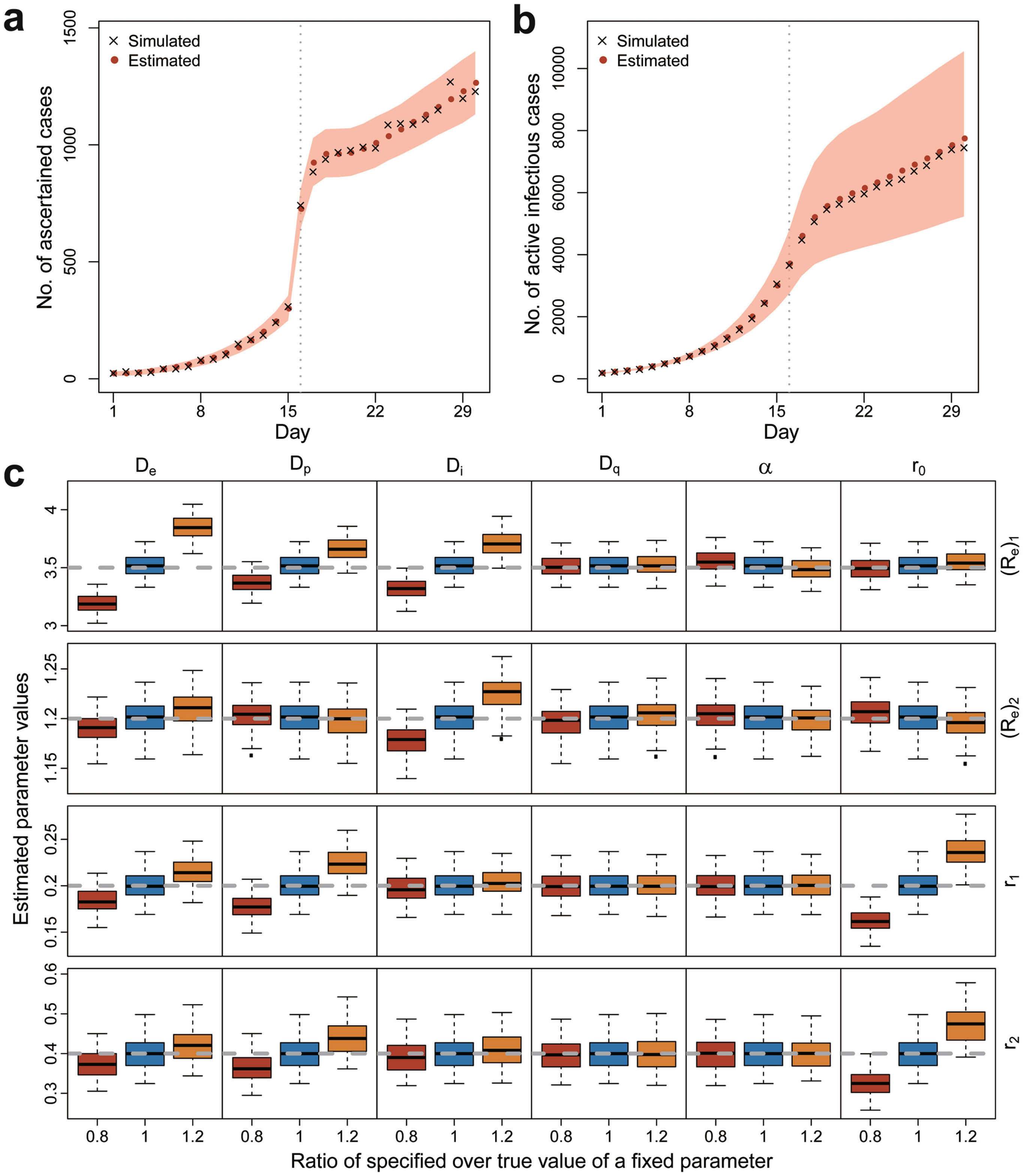
30. Haario, H., Laine, M., Mira, A. & Saksman, E. DRAM: efficient adaptive MCMC. *Stat. Comput.* **16**, 339–354 (2006).
31. Brooks, S. P. & Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**, 434–455 (1997).

Acknowledgements We thank H. Tian from Beijing Normal University for comments. This study was supported by the National Natural Scientific Foundation of China (91843302), the Fundamental Research Funds for the Central Universities (2019kfyXMBZ015), and the 111 Project (X.H., S.C., D.W., C.W., T.W.). X.L. is supported by Harvard University.

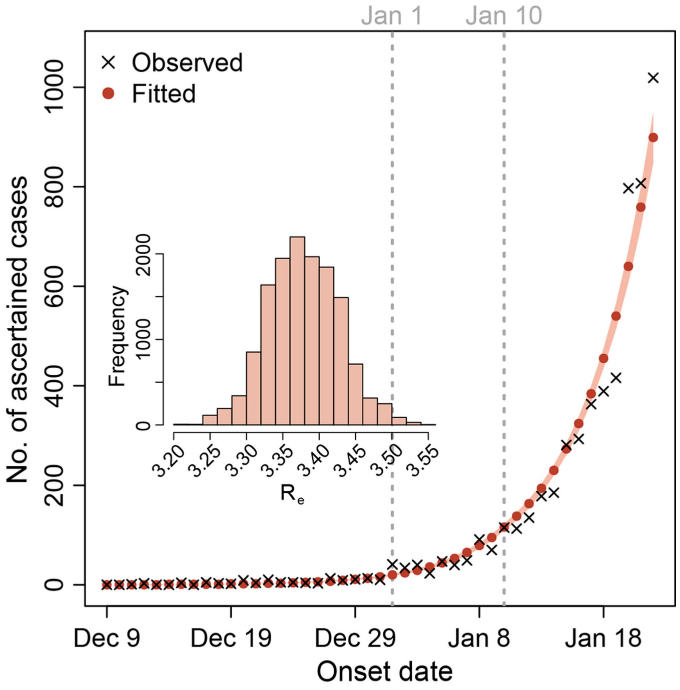
Author contributions T.W., X.L. and C.W. designed the study. X.H., S.C., X.L. and C.W. developed statistical methods. X.H., S.C. and D.W. performed data analysis. C.W. wrote the first draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2554-8.
Correspondence and requests for materials should be addressed to T.W., X.L. or C.W.
Peer review information *Nature* thanks David Fisman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Evaluation of the method on simulated data with two periods. **a**, **b**, Illustration of one simulated dataset. We estimated b_1, b_2, r_1 and r_2 when the other parameters were specified to their true values. The red points represent the mean estimates and the shaded areas indicate 95% credible intervals from 3,000 MCMC samples. **c**, Summary of results from 100 simulations. Each row represents an estimated parameter as indicated on the right, including $(R_e)_1$, $(R_e)_2$, r_1 and r_2. The grey dashed line in each row represents the true value of the parameter to be estimated. Each column represents a specified parameter as indicated on the top, including D_e, D_p, D_i, D_q, α and r_0, which we specified as the true values or 20% lower or higher than the true values. Each box summarizes estimates from 100 replicates, of which the median is indicated by the horizontal line, the interquartile range is indicated by the lower and upper bounds, and the minimum and the maximum are indicated by the whiskers.



Extended Data Fig. 2 | Estimation of R_0 using daily incidence data, starting from 9 December. Following the main analysis, we assumed $r_0 = 0.23$ and set $I(0) = 1$, $A(0) = 3$, $E(0) = 17$ and $P(0) = H(0) = R(0) = 0$ accordingly. We assumed transmission rate b, ascertainment rate r and duration from illness onset to hospitalization D_q (set to 21 days) were the same until 22 January 2020. All the other settings were the same as in the main analysis. The shaded area in the plot indicates 95% credible intervals estimated by the deterministic model with 10,000 sets of parameter values sampled from MCMC. Unlike other analyses, we did not construct 95% credible intervals by stochastic simulations, because stochastic variation at the early days had very large effects, owing to low counts. The inserted histogram shows the distribution of the estimated R_0 from 9 December 2019 to 22 January 2020, for which the mean estimate was 3.38 (95% credible interval 3.28–3.48).

Extended Data Table 1 | Notations of compartments and the initial values for the main analysis

Notation	Meaning	Initial value	Note
S	Number of susceptible individuals	9,999,021	$S = N - E - P - A - I - H - R$
E	Number of exposed cases	478	$E(0) = r_0^{-1} E_\mathrm{I}(0)$, where $E_\mathrm{I}(0)$ was the number of ascertained cases with onset during day $(D_\mathrm{p} + 1)$ and day $(D_\mathrm{p} + D_\mathrm{e})$ (Jan 3-5, 2020)*
P	Number of presymptomatic cases	326	$P(0) = r_0^{-1} P_\mathrm{I}(0)$, where $P_\mathrm{I}(0)$ was the number of ascertained cases with onset during day 1 and day D_p (Jan 1-2, 2020)*
I	Number of ascertained cases	34	Number of ascertained cases with onset within D_i days before day 1 (Dec 29-31, 2019)
A	Number of unascertained cases	114	$A(0) = r_0^{-1}(1 - r_0) I(0)$*
H	Number of isolated cases	27	Number of cases reported before day 1 (Jan 1, 2020)
R	Number of removed individuals	0	Number of cases recovered before day 1 (Jan 1, 2020)

*The initial ascertainment rate r_0 was assumed to be 0.23 in the main analysis. Day 1 is 1 January 2020.

Extended Data Table 2 | Parameter settings for five periods in the main analysis

Parameter	Meaning	Jan 1-9	Jan 10-22	Jan 23-Feb 1	Feb 2-16	Feb 17-Mar 8
b	Transmission rate of ascertained cases	b_{12}	b_{12}	b_3	b_4	b_5
r	Ascertainment rate	r_{12}	r_{12}	r_3	r_4	r_5
α	Ratio of transmission rate for unascertained over ascertained cases	0.55	0.55	0.55	0.55	0.55
D_e	Latent period	2.9	2.9	2.9	2.9	2.9
D_p	Presymptomatic infectious period	2.3	2.3	2.3	2.3	2.3
D_i	Symptomatic infectious period	2.9	2.9	2.9	2.9	2.9
D_q	Duration from illness onset to isolation	21	15	10	6	3
D_h	Isolation period	30	30	30	30	30
N	Population size	10,000,000	10,000,000	10,000,000	10,000,000	10,000,000
n	Daily inbound and outbound size	500,000	800,000	0	0	0

Extended Data Table 3 | COVID-19 cases exported from Wuhan to Singapore before 23 January 2020

Case ID	Arrival date	Symptom onset	Confirmed date
1	2020/1/20	2020/1/21	2020/1/23
2	2020/1/21	2020/1/21	2020/1/24
3	2020/1/20	2020/1/23	2020/1/24
4	2020/1/22	2020/1/23	2020/1/25
5	2020/1/18	2020/1/24	2020/1/27
6	2020/1/19	2020/1/25	2020/1/27
7	2020/1/23	2020/1/24	2020/1/27
8	2020/1/19	2020/1/24	2020/1/28
9	2020/1/19	2020/1/24	2020/1/29
10	2020/1/20	2020/1/21	2020/1/29
11	2020/1/22	2020/1/27	2020/1/29
12	2020/1/22	2020/1/26	2020/1/29
13	2020/1/21	2020/1/28	2020/1/30
16	2020/1/22	2020/1/23	2020/1/31
18	2020/1/22	2020/1/30	2020/2/1
26	2020/1/21	2020/1/28	2020/2/4

This information is from https://co.vid19.sg/singapore/dashboard.

Extended Data Table 4 | Estimated transmission rates from the main and sensitivity analyses

Analysis	Estimated transmission rates*				DIC†
	b_{12}	b_3	b_4	b_5	
Main	1.31 (1.25-1.37)	0.40 (0.38-0.43)	0.17 (0.16-0.19)	0.10 (0.08-0.12)	554.07
S1	1.31 (1.25-1.37)	0.37 (0.35-0.39)	0.17 (0.16-0.18)	0.10 (0.08-0.11)	387.63
S2	1.51 (1.43-1.57)	0.53 (0.51-0.56)	0.25 (0.24-0.27)	0.15 (0.13-0.17)	539.15
S3	1.46 (1.39-1.53)	0.34 (0.31-0.37)	0.11 (0.10-0.13)	0.04 (0.02-0.06)	588.73
S4	1.53 (1.46-1.61)	0.47 (0.44-0.50)	0.21 (0.19-0.22)	0.11 (0.09-0.13)	554.57
S5	1.18 (1.12-1.24)	0.36 (0.34-0.38)	0.16 (0.15-0.17)	0.09 (0.07-0.10)	553.49
S6	1.34 (1.28-1.39)	0.41 (0.38-0.44)	0.18 (0.17-0.19)	0.10 (0.08-0.12)	555.08
S7	1.27 (1.21-1.33)	0.39 (0.36-0.41)	0.17 (0.16-0.18)	0.10 (0.08-0.11)	555.40
S8	1.20 (1.14-1.27)	0.36 (0.34-0.39)	0.17 (0.16-0.18)	0.10 (0.08-0.12)	595.58
S9	0.93 (0.92-0.94)	0.26 (0.25-0.27)	0.17 (0.16-0.17)	0.16 (0.14-0.18)	808.38

*The estimates are displayed as mean (95% credible interval) based on 10,000 MCMC samples.

†The deviance information criterion (DIC) is presented for model comparison. Nevertheless, the DIC of S1 is not comparable to the others because the data of S1 were modified by smoothing the outlier data point on 1 February.

Extended Data Table 5 | Estimated R_e for different periods from the main and sensitivity analyses

Analysis	Estimated effective reproduction number R_e*				
	Jan 1-9	Jan 10-22	Jan 23-Feb 1	Feb 2-16	Feb 17-Mar 8
Main	3.54 (3.40-3.67)	3.32 (3.19-3.44)	1.18 (1.11-1.25)	0.51 (0.47-0.54)	0.28 (0.23-0.33)
S1	3.54 (3.40-3.67)	3.32 (3.19-3.44)	1.09 (1.02-1.16)	0.50 (0.47-0.54)	0.28 (0.23-0.32)
S2	3.21 (3.09-3.32)	3.03 (2.92-3.13)	1.23 (1.16-1.29)	0.57 (0.54-0.60)	0.33 (0.29-0.37)
S3	4.37 (4.19-4.55)	4.07 (3.91-4.24)	1.13 (1.04-1.22)	0.38 (0.34-0.41)	0.14 (0.08-0.19)
S4	3.56 (3.42-3.68)	3.34 (3.21-3.45)	1.18 (1.11-1.25)	0.51 (0.47-0.54)	0.27 (0.23-0.32)
S5	3.52 (3.39-3.66)	3.30 (3.18-3.43)	1.18 (1.11-1.25)	0.51 (0.47-0.54)	0.27 (0.23-0.32)
S6	3.52 (3.38-3.65)	3.29 (3.17-3.42)	1.19 (1.12-1.27)	0.51 (0.48-0.55)	0.28 (0.23-0.33)
S7	3.59 (3.46-3.72)	3.38 (3.26-3.49)	1.17 (1.10-1.24)	0.50 (0.47-0.53)	0.27 (0.23-0.32)
S8	3.79 (3.68-3.90)	3.58 (3.48-3.68)	1.15 (1.08-1.22)	0.50 (0.47-0.53)	0.27 (0.23-0.32)
S9	3.42 (3.40-3.45)	3.25 (3.23-3.27)	0.92 (0.88-0.95)	0.54 (0.51-0.56)	0.44 (0.38-0.49)

*The estimates are displayed as mean (95% credible interval) based on 10,000 MCMC samples.

Extended Data Table 6 | Estimated ascertainment rates from the main and sensitivity analyses

Analysis	Estimated ascertainment rate*				
	r_{12}	r_3	r_4	r_5	Overall
Main	0.15 (0.13-0.17)	0.14 (0.11-0.17)	0.10 (0.08-0.12)	0.16 (0.13-0.21)	0.13 (0.11-0.16)
S1	0.15 (0.12-0.17)	0.15 (0.12-0.18)	0.11 (0.09-0.14)	0.19 (0.14-0.24)	0.13 (0.11-0.16)
S2	0.14 (0.12-0.17)	0.15 (0.12-0.18)	0.10 (0.08-0.13)	0.17 (0.13-0.22)	0.14 (0.11-0.17)
S3	0.14 (0.12-0.16)	0.13 (0.10-0.16)	0.09 (0.07-0.11)	0.16 (0.12-0.20)	0.12 (0.10-0.15)
S4	0.15 (0.13-0.17)	0.14 (0.12-0.17)	0.10 (0.08-0.12)	0.17 (0.13-0.21)	0.13 (0.11-0.16)
S5	0.15 (0.13-0.17)	0.14 (0.11-0.17)	0.10 (0.08-0.12)	0.16 (0.12-0.21)	0.13 (0.11-0.16)
S6	0.09 (0.08-0.10)	0.09 (0.07-0.11)	0.06 (0.05-0.08)	0.10 (0.08-0.13)	0.08 (0.07-0.10)
S7	0.26 (0.22-0.30)	0.25 (0.20-0.30)	0.18 (0.14-0.22)	0.29 (0.22-0.37)	0.23 (0.16-0.28)
S8	0.55 (0.47-0.62)	0.50 (0.41-0.60)	0.35 (0.28-0.43)	0.59 (0.46-0.74)	0.47 (0.39-0.58)

*The estimates are displayed as mean (95% credible intervals) based on 10,000 MCMC samples.

Extended Data Table 7 | Prediction of the ending date of COVID-19 epidemic in Wuhan from the main and sensitivity analyses

Analysis	First day of no ascertained infections*	Clearance of all infections†
Main	Mar 27 (Mar 20 to Apr 5)‡	Apr 21 (Apr 8 to May 12)
S1	Mar 27 (Mar 20 to Apr 4)	Apr 20 (Apr 7 to May 11)
S2	Mar 28 (Mar 21 to Apr 5)	Apr 22 (Apr 8 to May 13)
S3	Mar 25 (Mar 18 to Apr 2)	Apr 19 (Apr 5 to May 8)
S4	Mar 27 (Mar 20 to Apr 4)	Apr 21 (Apr 8 to May 12)
S5	Mar 27 (Mar 20 to Apr 4)	Apr 21 (Apr 8 to May 13)
S6	Mar 27 (Mar 20 to Apr 4)	Apr 24 (Apr 11 to May 15)
S7	Mar 27 (Mar 20 to Apr 4)	Apr 17 (Apr 4 to May 7)
S8	Mar 26 (Mar 19 to Apr 4)	Apr 10 (Mar 29 to Apr 30)
S9	Apr 5 (Mar 26 to Apr 18)	Apr 20 (Apr 4 to May 16)

*First day of no ascertained infections means the first day of $I = 0$.

†Clearance of all infections means the first day of $E = P = A = I = 0$.

‡The estimates are displayed as mean date (95% credible interval) based on 10,000 stochastic simulations with parameter values from MCMC sampling.

nature research

Corresponding author(s): Chaolong Wang, Xihong Lin, Tangchun Wu

Last updated by author(s): Jun 27, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SAS version 9.4 was used for data collection.
Data analysis	Data analysis was performed in R (version 3.6.2), alongside with third-party R packages BayesianTools (version 0.1.7) and coda (version 0.19.3). R codes are available on Github via link: https://github.com/chaolongwang/SAPHIRE.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data are available on Github via link: https://github.com/chaolongwang/SAPHIRE.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study contained 32,583 laboratory-confirmed COVID-19 cases between 18 Dec 2019 and 8 Mar 2020 in Wuhan. We included all laboratory-confirmed COVID-19 cases that were reported by 8 Mar 2020.
Data exclusions	For the consistency of case definition throughout different time periods, we excluded COVID-19 cases diagnosed by clinical symptoms without SARS-CoV-2 virus confirmation by the real-time reverse-transcription-polymerase-chain-reaction (RT-PCR) assay or high-throughput sequencing of nasal and pharyngeal swab specimens. Exclusion criterion was pre-established.
Replication	Not applicable because this study is retrospective and observational.
Randomization	Not applicable because this study is retrospective and observational.
Blinding	Not applicable because this study is retrospective and observational.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo’

https://doi.org/10.1038/s41586-020-2488-1

Received: 2 April 2020

Accepted: 23 June 2020

Enrico Lavezzo[1,13], Elisa Franchin[1,13], Constanze Ciavarella[2], Gina Cuomo-Dannenburg[2], Luisa Barzon[1], Claudia Del Vecchio[1], Lucia Rossi[3], Riccardo Manganelli[1], Arianna Loregian[1], Nicolò Navarin[4,5], Davide Abate[1], Manuela Sciro[3], Stefano Merigliano[6], Ettore De Canale[3], Maria Cristina Vanuzzo[3], Valeria Besutti[3], Francesca Saluzzo[1], Francesco Onelia[1], Monia Pacenti[3], Saverio G. Parisi[1], Giovanni Carretta[3], Daniele Donato[3], Luciano Flor[3], Silvia Cocchio[7], Giulia Masi[1], Alessandro Sperduti[4,5], Lorenzo Cattarino[2], Renato Salvador[6], Michele Nicoletti[8], Federico Caldart[8], Gioele Castelli[8], Eleonora Nieddu[8], Beatrice Labella[8], Ludovico Fava[8], Matteo Drigo[8], Katy A. M. Gaythorpe[2], Imperial College COVID-19 Response Team*, Alessandra R. Brazzale[9], Stefano Toppo[1,5], Marta Trevisan[1], Vincenzo Baldo[7], Christl A. Donnelly[2,10], Neil M. Ferguson[2], Ilaria Dorigatti[2,14] ✉ & Andrea Crisanti[1,11,14] ✉

On 21 February 2020, a resident of the municipality of Vo’, a small town near Padua (Italy), died of pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection[1]. This was the first coronavirus disease 19 (COVID-19)-related death detected in Italy since the detection of SARS-CoV-2 in the Chinese city of Wuhan, Hubei province[2]. In response, the regional authorities imposed the lockdown of the whole municipality for 14 days[3]. Here we collected information on the demography, clinical presentation, hospitalization, contact network and the presence of SARS-CoV-2 infection in nasopharyngeal swabs for 85.9% and 71.5% of the population of Vo’ at two consecutive time points. From the first survey, which was conducted around the time the town lockdown started, we found a prevalence of infection of 2.6% (95% confidence interval (CI): 2.1–3.3%). From the second survey, which was conducted at the end of the lockdown, we found a prevalence of 1.2% (95% CI: 0.8–1.8%). Notably, 42.5% (95% CI: 31.5–54.6%) of the confirmed SARS-CoV-2 infections detected across the two surveys were asymptomatic (that is, did not have symptoms at the time of swab testing and did not develop symptoms afterwards). The mean serial interval was 7.2 days (95% CI: 5.9–9.6). We found no statistically significant difference in the viral load of symptomatic versus asymptomatic infections ($P = 0.62$ and 0.74 for *E* and *RdRp* genes, respectively, exact Wilcoxon–Mann–Whitney test). This study sheds light on the frequency of asymptomatic SARS-CoV-2 infection, their infectivity (as measured by the viral load) and provides insights into its transmission dynamics and the efficacy of the implemented control measures.

As of 23 May 2020, 5,105,881 confirmed cases and 333,446 deaths of COVID-19 have been reported worldwide[2]. In Italy, COVID-19 has caused more than 32,616 confirmed deaths. The causative agent (SARS-CoV-2), a close relative of SARS-CoV[4], was detected in the human population in Wuhan city, Hubei province (China) around the beginning of December 2019[5,6]. In Hubei province and in the rest of mainland China, recent reports suggest that strategies based on the isolation of cases and their contacts, along with drastic social distancing measures that include the quarantine of whole cities and regions, the closure of schools and workplaces and the cancellations of mass gatherings had a considerable effect on the control of the epidemic[7,8]. However, the long-term effectiveness of these interventions remains unclear[9]. In Europe, similar interventions have been implemented to control the transmission of SARS-CoV-2. Recent analyses suggest that control is likely to be achieved across Europe[10]. In Italy, interventions have successfully controlled the transmission of SARS-CoV-2 in all regions, but uncertainties remain about the ability to avoid a resurgence of transmission as interventions are relaxed[11]. Effective long-term control of transmission in Europe and worldwide depends on an improved understanding of the mechanisms of SARS-CoV-2 transmission, particularly on the relative contribution

---

[1]Department of Molecular Medicine, University of Padova, Padua, Italy. [2]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK. [3]Azienda Ospedale Padova, Padua, Italy. [4]Department of Mathematics "Tullio Levi-Civita", University of Padova, Padua, Italy. [5]CRIBI Biotech Center, University of Padova, Padua, Italy. [6]Department of Surgery, Oncology and Gastroenterology, University of Padova, Padua, Italy. [7]Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, Padua, Italy. [8]School of Medicine, University of Padova, Padua, Italy. [9]Department of Statistical Sciences, University of Padova, Padua, Italy. [10]Department of Statistics, University of Oxford, Oxford, UK. [11]Department of Life Sciences, Imperial College London, London, UK. [13]These authors contributed equally: Enrico Lavezzo, Elisa Franchin. [14]These authors jointly supervised this work: Ilaria Dorigatti, Andrea Crisanti. *A list of authors and their affiliation appears at the end of the paper. ✉e-mail: i.dorigatti@imperial.ac.uk; andrea.crisanti@unipd.it

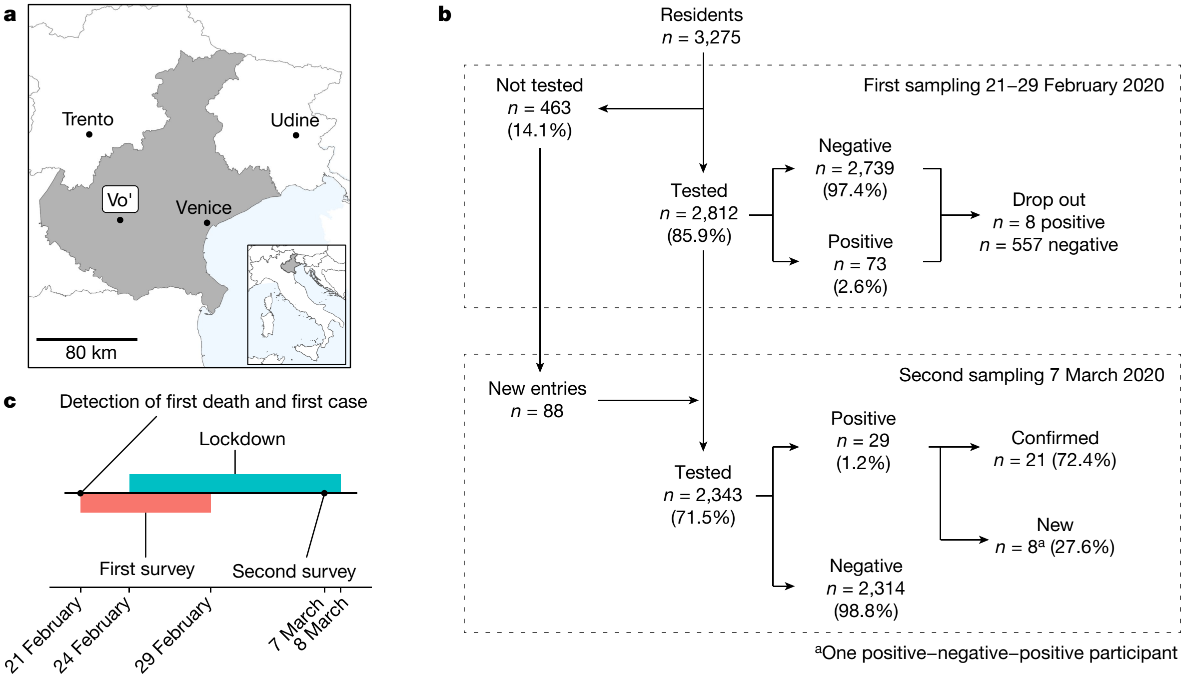


Fig. 1 | Study description. a, Map showing the location of Vo' and the Veneto region (grey area) within Italy, produced using shapefiles from GADM (https://gadm.org/) and Italian National Institute of Statistics (ISTAT; https://www.istat.it/it/archivio/222527 and https://www.istat.it/it/archivio/104317#accordions). **b**, Flow chart summarizing the key statistics on the two sequential nasopharyngeal swab surveys conducted in Vo' to assess the transmission of SARS-CoV-2 before and after the implementation of interventions. **c**, Summary of the key events in the study period.

of asymptomatic, presymptomatic and symptomatic transmission[12]. This is particularly important given that, in the absence of a vaccine or effective treatment, alternative public health interventions are being evaluated to allow the population to maintain essential societal and economic activities, while controlling the spread of SARS-CoV-2, limiting mortality and maintaining healthcare demand within capacity.

In this study, we present the results of two surveys of the resident population of Vo', conducted less than 2 weeks apart, to investigate population exposure to SARS-CoV-2 before and after the lockdown. We present an analysis of population demography, the prevalence of infection, viral load and the frequency of symptomatic, asymptomatic and presymptomatic infections. We assessed the risk of SARS-CoV-2 infection associated with comorbidity and therapies for underlying conditions, characterized chains of transmission, studied the transmission dynamics of SARS-CoV-2 and assessed the impact of the lockdown. Our analyses show that viral transmission could be effectively and rapidly suppressed by combining the early isolation of infected people with community lockdown. The experience of Vo' shows that, despite the silent and widespread transmission of SARS-CoV-2, transmission can be controlled and represents a model for the systematic suppression of viral outbreaks under similar epidemiological and demographic conditions.

During the two surveys, we collected nasopharyngeal swabs from 2,812 and 2,343 study participants, which corresponded to 85.9% and 71.5% of the eligible study population, respectively (Fig. 1). All age groups were homogeneously sampled with age-specific percentages ranging from 57.1% to 95.4% in the first survey and 40.1% to 80.4% in the second survey (Extended Data Table 1). Statistical analysis showed that, while the recruited and non-recruited populations are different in terms of age distribution ($P < 0.001$ for the first and second surveys, Fisher's exact test), there was no statistically significant bias in the composition of the different age groups enrolled in the two surveys ($P = 0.31$, exact Wilcoxon–Mann–Whitney test) (Extended Data Fig. 1). Notably, no additional infections were reported in Vo' despite the escalating epidemic in the surrounding regions.

Analysis of infection prevalence

A total of 73 out of the 2,812 participants who were tested at the first survey were positive, which gives a prevalence of 2.6% (95% CI: 2.1–3.3%) (Table 1). The second survey identified 29 total positive cases (prevalence of 1.2%; 95% CI: 0.8–1.8%), 8 of which were new cases (prevalence of 0.3%; 95% CI: 0.15–0.7%) (Fig. 2). One of the eight new infections detected in the second survey was a hospitalized participant who tested positive, then negative, then positive again. It is unclear whether this was a case of SARS-CoV-2 re-infection or whether the second test was a false negative. The frequency of the symptoms in the participants who were positive for SARS-CoV-2 infection was systematically recorded, with fever and cough being the most common (Extended Data Fig. 1). Notably, a total of 29 out of the 73 participants (39.7%; 95% CI: 28.5–51.9%) who tested positive at the first survey were asymptomatic (that is, did not show symptoms at the time of swab sampling nor afterwards; see the definition of symptomatic in the Methods). A similar

Table 1 | Participants positive for SARS-CoV-2 at the first and second surveys

	First survey		Second survey	
	Total positives	**Percentage**	**Total positives**	**Percentage**
Symptomatic at the time of sampling[a]	34	46.6	15	51.7
Presymptomatic at the time of sampling	10	13.7	1	3.4
Asymptomatic	29	39.7	13	44.8
Total	73		29	

[a] Defined as the presence of hospitalization and/or fever and/or cough and/or at least two of the following symptoms: sore throat, headache, diarrhoea, vomit, asthenia, muscle pain, joint pain, and loss of taste or smell.

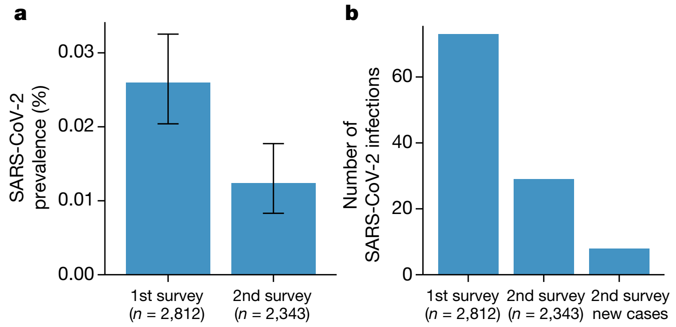


Fig. 2 | SARS-CoV-2 prevalence statistics. a, The prevalence of SARS-CoV-2 infection at the first survey ($x = 73$ positive out of $n = 2{,}812$ tested) and the second survey ($x = 29$ positive out of $n = 2{,}343$ tested). The error bars represent the 95% exact binomial CI. **b**, The number of SARS-CoV-2 infections detected in the sampled population of the residents of Vo' in the first survey ($x = 73$) and the second survey ($x = 29$, of which 8 were new infections).

proportion of asymptomatic infection was also recorded at the second survey (13 out of 29, 44.8%; 95% CI: 26.5–64.3%); of the eight new cases, five were asymptomatic (Table 2, Extended Data Fig. 2). No infections were detected in either survey in 234 tested children ranging from 0 to 10 years of age, including those living in the same household as infected individuals (Extended Data Table 3). The prevalence of infection oscillated between a central estimate of 1.2% and 1.7% up to 50 years of age (Extended Data Fig. 1). Older participants showed a threefold increase in the infection prevalence (Table 2, Extended Data Fig. 1). Of the 81 participants who were positive for SARS-CoV-2 across the two surveys, 13 required hospitalization (16.0%). Their age distribution was as follows: 1 (7.7%) aged 41–50 years, 1 (7.7%) aged 51–60 years, 4 (30.8%) aged 61–70 years, 5 (38.5%) aged 71–80 years and 2 (15.4%) aged 81–90 years.

A substantial fraction of infected participants (58.9%; 95% CI: 46.8–70.3%, presymptomatic, symptomatic and asymptomatic combined over all ages) cleared the infection between the first and second surveys, that is, had a negative test at the second survey after a positive test at the first survey (Extended Data Table 2). For all infections identified in the study, clearance was confirmed by an additional negative test that was conducted independently by the local health authority (data not shown). The time to viral clearance (the time from the earliest positive sample for the participants with more than one sample in the first survey to a negative sample in the second survey) ranged from 8 to 13 days and was on average 9.3 days, with a standard deviation of 2.0 days. The minimal duration of the positivity window (the time from the earliest positive sample in the first survey to a positive sample in the second survey) ranged from 3 to 13 days and was on average 9.1 days, with a standard deviation of 2.3 days. In particular, 61.4% (95% CI: 45.5–75.6%) of symptomatic and 55.2% (95% CI: 35.7–73.6%) of asymptomatic individuals with SARS-CoV-2 infections cleared the virus during the study period (that is, had a negative test after a positive result at the first survey); the highest rate (100%) was observed in the age groups of symptomatic individuals aged 31–40 and 41–50 years (Extended Data Table 2). SARS-CoV-2 positivity overall (that is, the first and second surveys combined) and at the first survey was more frequently associated with individuals who were 71–80 years of age (compared to those 21–30 years of age; $P = 0.012$ and $P = 0.017$, respectively) (Extended Data Fig. 1). Being male was associated with SARS-CoV-2 positivity in the second survey ($P = 0.04$) (Table 2). Analyses of the association between common comorbidities such as diabetes, hypertension, vascular diseases, respiratory diseases in asymptomatic and symptomatic people and the use of treatment for a number of different conditions with symptomatic infection showed no significant association (Supplementary Tables 3, 4).

Role of asymptomatic transmission

The analysis of viral genome equivalents inferred from cycle threshold data from real-time reverse transcription PCR (RT–PCR) assays indicated that asymptomatic and symptomatic participants did not differ when data from viral PCR templates recovered from the nasopharyngeal swabs of asymptomatic and symptomatic participants were compared ($P = 0.62$ and 0.74 for gene *E* and gene *RdRp*, respectively; exact Wilcoxon–Mann–Whitney test) (Extended Data Fig. 3). We found that the viral load tends to peak around the day of symptom onset and, for most of the participants, tends to decline after

Table 2 | Participants testing positive stratified by sex and age

	First survey			Second survey				
	n	Positive	Percentage	*n*	Positive	Percentage	New positive	Percentage
Sex								
Males	1,408	43	3.1	1,165	20	1.7	5	0.4
Females	1,404	30	2.1	1,178	9	0.8	3	0.3
P value			0.15			0.041		
Age group								
0–10	217	0	0.0	157	0	0.0	0	0.0
11–20	250	3	1.2	210	2	1.0	1	0.5
21–30	240	4	1.7	191	2	1.0	0	0.0
31–40	286	7	2.4	241	2	0.8	0	0.0
41–50	439	5	1.1	366	2	0.5	1	0.3
51–60	496	16	3.2	439	7	1.6	2	0.5
61–70	384	15	3.9	349	6	1.7	2	0.6
71–80	318	19	6.0	262	6	2.3	2	0.8
>81	182	4	2.2	128	2	1.6	0	0.0
P value			<0.001[a]			0.48		
Total	2,812	73	2.6	2,343	29	1.2	8	0.3

P values (two-sided) were computed using Fisher's exact test (for sex) and the likelihood ratio test (for age group).

[a] Linear trend.

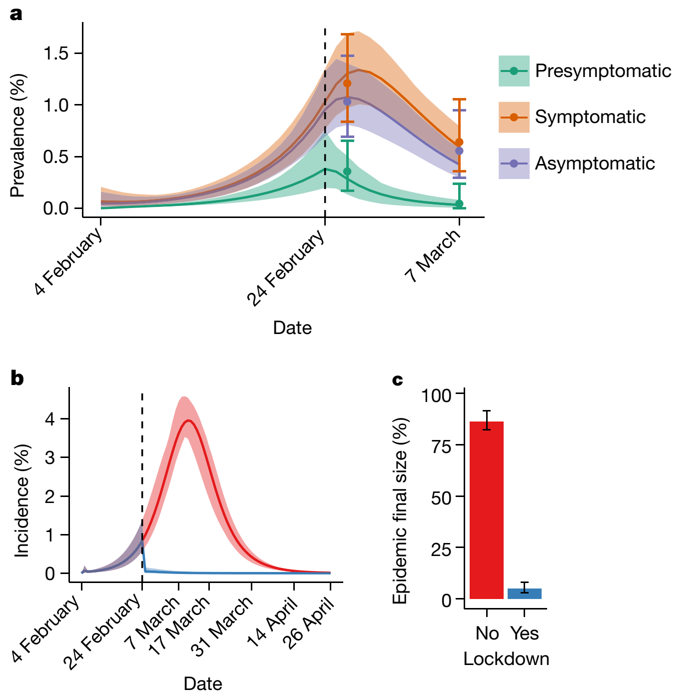


Fig. 3 | SARS-CoV-2 dynamics of the mitigated and counterfactual unmitigated epidemic in Vo' and the relative final size estimates. a, The prevalence of SARS-CoV-2 infection inferred from the observed prevalence data for symptomatic, presymptomatic and asymptomatic infections in the first and second surveys using R_0^1 (the reproduction number before the lockdown) = 2.4 and $1/\sigma$ (the average duration of positivity beyond the duration of the infectious period) = 4 days. The dashed vertical line represents the time that the lockdown started. The points represent the observed prevalence data, the 95% CI is the exact binomial CI. The solid lines represent the mean and the shading represents the 95% credible interval obtained with the model from 100 samples from the posterior distribution of the parameters. **b**, The incidence of the epidemic fitted to the prevalence data (blue) and of the unmitigated epidemic (red), obtained assuming the same initial reproduction number value $R_0^1 = 2.4$ throughout the whole epidemic and $1/\sigma = 4$ days. The dashed vertical line represents the time that the lockdown started. The solid lines represent the mean and the shading represents the 95% credible interval obtained with the model from 100 samples from the posterior distribution of the parameters. **c**, The mean epidemic final size (the proportion of the population infected at the end of the epidemic) of the counterfactual unmitigated epidemic (red) and of the epidemic fitted from the prevalence data with the lockdown (blue). The error bars represent the range (minimum to maximum) of the mean final size obtained from $n = 100$ independent samples drawn from the posterior distribution of the parameters, calculated over the models with DIC (deviance information criterion) < 36.4.

symptom onset (Extended Data Fig. 3). The relative risk of contracting the infection from having close contacts with an infected relative, including those living in the same household, gives an odds ratio of 84.5 (95% CI: 16.8–425.4) (Extended Data Table 4, Supplementary Text 3). Two out of the eight participants with new infections that were detected in the second survey either shared a household or had a contact history with asymptomatic individuals (Supplementary Table 1).

Reconstructing transmission chains

From the inferred transmission pairs, we estimated a serial interval distribution over the whole study period with a mean of 7.2 days (95% CI: 5.9–9.6). We found that the lockdown reduced the serial interval from a mean of 7.6 days (95% CI: 6.4–8.7) before the lockdown to a mean of 6.2 days (95% CI: 2.6–10.7) after the lockdown (Extended Data Fig. 4). We also found that the lockdown substantially reduced transmission, with the reproduction number dropping from an initial value of 2.49 (95% CI: 1.31–4.00) before the lockdown to 0.41 (95% CI: 0.21–0.63) after the lockdown.

Modelling point prevalence data

We used the prevalence estimates obtained in Vo' at the first and second surveys to calibrate a modified susceptible–exposed–infectious–recovered compartmental model of SARS-CoV-2 transmission that incorporates symptomatic, presymptomatic and asymptomatic infections, virus detectability (in swabs) before and after the infectious period and the impact of the lockdown (Extended Data Fig. 5). We assumed that presymptomatic, symptomatic and asymptomatic infections transmit the virus. We estimated that on average 41% of the infections are asymptomatic, that the mean infectious period is approximately 3.6–6.5 days, and that the lockdown reduced SARS-CoV-2 transmissibility on average by between 82% and 98%, depending on the assumed initial value of R_0^1 and on the duration of virus detectability (Supplementary Table 5). The model suggests that on average up to 86.2% (range: 82.2–91.6%) of the population would have been infected in the absence of interventions and that with the lockdown, 4.9% (range: 2.9–8.1%) of the population of Vo' was infected by SARS-CoV-2 (Fig. 3). These estimates are in line with the attack rates that were recently estimated for the Veneto region[11]. The model suggests that shorter values of the average duration of virus detectability beyond the infectious period better capture the central point prevalence estimates (Extended Data Fig. 6, Supplementary Table 5). Our results suggest that SARS-CoV-2 was introduced into the Vo' population at the beginning of February 2020.

Discussion

The results of the two surveys carried out in Vo' provide important insights into the transmission dynamics of SARS-CoV-2. Our finding that 42.5% (95% CI: 31.5–54.6%) of all confirmed SARS-CoV-2 infections across the two surveys were asymptomatic is in accordance with other population surveys[13]. Among confirmed SARS-CoV-2 infections, we did not observe significant differences in the frequency of asymptomatic infection between age groups (Supplementary Fig. 10; $P = 0.96$, Fisher's exact test). Among symptomatic participants, older age groups tended to show higher frequencies of SARS-CoV-2 infection (Extended Data Table 2). Recent studies have found that the clinical progression of infection in children is generally milder than in adults[14–16]. We found that none of the children under 10 years of age who took part in the study tested positive for SARS-CoV-2 infection at either survey, despite at least 13 of them living together with infected family members (Extended Data Table 3). This agrees with a recent study conducted in Iceland[13] and is particularly intriguing given the very high observed odds ratio for adults to become infected when living together with family members who are positive for SARS-CoV-2. However, this result does not mean that children cannot be infected by SARS-CoV-2, but suggests that children may be less susceptible than adults. The pathogenesis of SARS-CoV-2 infection in young children is not well understood[15]. Notably, nasopharyngeal swabs are tested for the presence of SARS-CoV-2 and can only detect active infection, not exposure. A cross-sectional serological survey would clarify the actual infection rates of the whole population, including children's exposure, to SARS-CoV-2.

The contribution of asymptomatic infections to SARS-CoV-2 transmission is supported by the viral load data (Extended Data Fig. 3), by the model fit to the observed prevalence data (Extended Data Fig. 6, Supplementary Table 5) and by the observation that two out of the eight participants with new infections that were detected in the second survey reported contacts with asymptomatic individuals (Supplementary Text 3). The extent to which symptoms may promote viral shedding remains to be determined, but the decreasing trend in viral load post-symptom onset suggests that presymptomatic transmission may play an important part[17]. Asymptomatic transmission and presymptomatic transmission pose clear challenges for the control of COVID-19 in the absence of strict social distancing measures or active epidemiological surveillance comprising, for instance, a test, trace and isolate strategy.

This study has informed the policy adopted by the Veneto region, where swabs are available to all contacts of positive symptomatic cases. This testing and tracing approach has had a tremendous impact on the course of the epidemic in Veneto compared to other Italian regions. In this context, the control strategy applied to Vo' serves as a model to suppress SARS-CoV-2 transmission across spatial scales. Enhanced community surveillance, the early detection of SARS-CoV-2 transmission and the timely implementation of interventions are key to control COVID-19 and reduce its substantial public health, economic and societal burden worldwide.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2488-1.

1. Crisanti, A. & Cassone, A. In one Italian town, we showed mass testing could eradicate the coronavirus. *The Guardian* https://www.theguardian.com/commentisfree/2020/mar/20/eradicated-coronavirus-mass-testing-covid-19-italy-vo (2020).
2. World Health Organization. Novel coronavirus (COVID-19) situation. *WHO* https://covid19.who.int/ (2020).
3. Saini, V. Coronavirus: voices from a quarantined Italian town. *EU Observer* https://euobserver.com/coronavirus/147552 (2020).
4. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
5. Volz, E. et al. Report 5: phylogenetic analysis of SARS-CoV-2. *Imperial College London* https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-02-15-COVID19-Report-5.pdf (2020).
6. Centre for Health Protection of the Hong Kong Special Administrative Region Government. CHP closely monitors cluster of pneumonia cases on Mainland. *The Government of the Hong Kong Special Administrative Region* https://www.info.gov.hk/gia/general/201912/31/P2019123100667.htm (2020).
7. Prem, K. et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).
8. Lai, S. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* https://doi.org/10.1038/s41586-020-2293-x (2020).
9. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **395**, 931–934 (2020).
10. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* https://doi.org/10.1038/s41586-020-2405-7 (2020).
11. Vollmer, M. A. C. et al. Report 20: using mobility to estimate the transmission intensity of COVID-19 in Italy: a subnational analysis with future scenarios. *Imperial College London* https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-05-04-COVID19-Report-20.pdf (2020).
12. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill*. **25**, 200018 (2020).
13. Gudbjartsson, D. F. et al. Spread of SARS-CoV-2 in the Icelandic population. *N. Engl. J. Med*. **382**, 2302–2315 (2020).
14. Cai, J. et al. A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. *Clin. Infect. Dis*. https://doi.org/10.1093/cid/ciaa198 (2020).
15. Zimmermann, P. & Curtis, N. Coronavirus infections in children including COVID-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children. *Pediatr. Infect. Dis. J*. **39**, 355–368 (2020).
16. Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis*. https://doi.org/10.1016/S1473-3099(20)30287-5 (2020).
17. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med*. **26**, 672–675 (2020).

Imperial College COVID-19 Response Team

Kylie E. C. Ainslie[2], Marc Baguelin[2], Samir Bhatt[2], Adhiratha Boonyasiri[2], Olivia Boyd[2], Lorenzo Cattarino[2], Constanze Ciavarella[2], Helen L. Coupland[2], Zulma Cucunubá[2], Gina Cuomo-Dannenburg[2], Bimandra A. Djafaara[2], Christl A. Donnelly[2,10], Ilaria Dorigatti[2], Sabine L. van Elsland[2], Rich FitzJohn[2], Seth Flaxman[12], Katy A. M. Gaythorpe[2], Will D. Green[2], Timothy Hallett[2], Arran Hamlet[2], David Haw[2], Natsuko Imai[2], Benjamin Jeffrey[2], Edward Knock[2], Daniel J. Laydon[2], Thomas Mellan[2], Swapnil Mishra[2], Gemma Nedjati-Gilani[2], Pierre Nouvellet[2], Lucy C. Okell[2], Kris V. Parag[2], Steven Riley[2], Hayley A. Thompson[2], H. Juliette T. Unwin[2], Robert Verity[2], Michaela A. C. Vollmer[2], Patrick G. T. Walker[2], Caroline E. Walters[2], Haowei Wang[2], Yuanrong Wang[2], Oliver J. Watson[2], Charles Whittaker[2], Lilith K. Whittles[2], Xiaoyue Xi[2] & Neil M. Ferguson[2]

[12] Department of Mathematics, Imperial College London, London, UK.

Methods

Study setting

The municipality of Vo', in the province of Padua, Veneto region, Italy, is located about 50 km west of Venice (Fig. 1a). According to the latest land registry, Vo' has a population of 3,275 individuals over an area of 20.4 km^2. Upon the detection of SARS-CoV-2 in a deceased resident of Vo' on 21 February 2020, the same day where the first COVID-19 case was detected in Vo' and 1 day after the first locally acquired COVID-19 infection was identified in Italy, we conducted an epidemiological study to investigate the prevalence of SARS-CoV-2 infection in the population. Sampling was conducted on the majority of the Vo' population at two time points: the first during the days immediately after the detection of the first cases (21–29 February 2020), and the second one at the end of the 2-week lockdown (7 March 2020) (Fig. 1c). For each resident, we collected information on the sampling dates, the results of SARS-CoV-2 testing, demographics (for example, age and sex), residence, health record (including symptoms and COVID-19-related hospitalization dates, previous conditions and therapy taken for other illnesses), household size and contact network. The data were collated using Microsoft Excel and the data set spreadsheet is available at https://github.com/ncov-ic/SEIR_Covid_Vo. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

The definition of symptomatic is as follows: a participant who required hospitalization and/or reported fever (yes/no or a temperature above 37 °C) and/or cough and/or at least two of the following symptoms: sore throat, headache, diarrhoea, vomit, asthenia, muscle pain, joint pain, loss of taste or smell, or shortness of breath.

Laboratory methods

Upper respiratory tract samples were collected by healthcare professionals with a single flocked tapered swab used for the oropharynx then nasal mid-turbinates and immediately put into a sterile tube containing transport medium (eSwab, Copan Italia Spa). Sampling was performed according the US Centers for Disease Control and Prevention guidelines[18]. In brief, for oropharyngeal sampling, the swab was inserted into the posterior pharynx and tonsillar areas and rubbed over both tonsillar pillars and posterior oropharynx, avoiding touching the tongue, teeth and gums; for deep nasal sampling, the swab was inserted into both nostrils for about 2 cm while gently rotating against the nasal wall several times. Samples were stored at 2–8 °C until testing, which was performed within 72 h from collection. As a measure of the correct execution of the sampling, each PCR contains an internal control designed to amplify the human housekeeping gene encoding RNase P. Reactions that failed to show the internal positive control were classified as invalid and repeated. Total nucleic acids were purified from 200 μl of nasopharyngeal swab samples and eluted in a final volume of 100 μl by using a MagNA Pure 96 System (Roche Applied Sciences). Detection of SARS-CoV-2 RNA was performed by an in-house real-time RT–PCR method, which was developed according the protocol and the primers and probes designed by Corman et al.[19] that targeted the genes encoding envelope (*E*) (E_Sarbeco_F, E_Sarbeco_R and E_Sarbeco_P1) and RNA-dependent RNA polymerase (*RdRp*: RdRp_SARSr-F, RdRp_SARSr-R, RdRP_SARSr-P1 and RdRp_SARSr-P2) of SARS-CoV-2. Real-time RT–PCR assays were performed in a final volume of 25 μl, containing 5 μl of purified nucleic acids, using One Step Real Time kit (Thermo Fisher Scientific) and run on ABI 7900HT Fast Sequence Detection Systems (Thermo Fisher Scientific). The sensitivity of the *E* and *RdRp* gene assays was 5.0 and 50 genome equivalent copies per reaction at 95% detection probability, respectively. Both assays had no cross-reactivity with the endemic human coronaviruses HCoV-229E, HCoV-NL63, HCoV-OC43 and HCoV-HKU1 and with MERS-CoV. All tests were performed at the Clinical Microbiology and Virology Unit of Padova University Hospital, which is the Regional Reference Laboratory for emerging viral infections. After an initial period of dual testing by the National Reference Laboratory at the Italian Institute of Health (Istituto Superiore di Sanità), which demonstrated 100% agreement of results, the Regional Reference Laboratory received accreditation as Reference Laboratory for COVID-19 testing.

Assessment of genome equivalents

Cycle threshold (C_t) data from real-time RT–PCR assays were collected for *E* and *RdRp* genes. C_t data for the *E* gene were available for 30 symptomatic, 5 presymptomatic and 23 asymptomatic infections, and for the *RdRp* gene for 27 symptomatic, 9 presymptomatic and 26 asymptomatic infections. Genome equivalent copies per ml were inferred according to linear regression performed on calibration standard curves. The interpolated C_t values were further multiplied by 100, according to the final dilution factor (1:100). Linear regression was calculated in Python3.7.3 using modules scipy 1.4.1, numpy 1.18.1 and matplotlib 3.2.1[20]. Genome equivalent distributions from the two genes, for positive symptomatic, asymptomatic and presymptomatic participants were compared with the exact Wilcoxon–Mann–Whitney test. Both viral load genome equivalents and raw C_t data are provided in the data set.

Reconstructing transmission chains

We used data on contacts traced within the community and on household contacts derived from household composition data (reported in the data set) to impute chains of transmission and transmission clusters. We used the R package epicontacts[21,22] to visualize the reconstructed transmission chains. We provide the algorithms used to infer the serial interval (the time from symptom onset of the infector to symptom onset of the infectee) distribution and the effective reproduction number (the average number of secondary infections generated by the identified infectors) in Supplementary Information Text 1 and 2, respectively. In brief, we inferred the date of symptom onset for the participants who tested positive but with a missing onset date from the observed time-lags from symptom onset to confirmation (for the participants who tested positive at multiple sampling times, we used the first sampling time). We then used the observed and inferred dates of symptom onset alongside the contact information to infer transmission pairs within the sampled population. In turn, reconstructed transmission pairs were used to characterize the serial interval in the whole study period as well as during the pre-lockdown and post-lockdown periods. Central effective reproduction number estimates were calculated as the average number of secondary infections generated by observed or imputed infectors, having assigned the infector stochastically when more than one or no potential infectors were identified. The 95% CIs were estimated by bootstrapping. All details are provided in Supplementary Information Text 1 and 2.

Mathematical modelling

The first survey occurred between 21 and 29 February 2020 and the second survey occurred on 7 March 2020. In the model, we assumed that prevalence was taken on the weighted average of the first sample collection date, that is, on 26 February 2020 and on 7 March 2020. The flow diagram of the model is given in Extended Data Fig. 5. We assumed that the population of Vo' was fully susceptible to SARS-CoV-2 (S compartment) at the start of the epidemic. Upon infection, infected people incubate the virus (E compartment) and have undetectable viraemia for an average of $1/\nu$ days before entering a stage (TP compartment) that lasts an average of $1/\delta$ days, in which people show no symptoms and have detectable viraemia. We assume that a proportion p of the infected population remains asymptomatic throughout the whole course of the infection (I_A compartment) and that the remaining proportion $1-p$ develops symptoms (I_S compartment). We assume that symptomatic (I_S), asymptomatic ($I_A + p\mathrm{TP}$) and presymptomatic ($(1-p)\mathrm{TP}$) people contribute to the onward transmission of SARS-CoV-2 and that symptomatic, asymptomatic and presymptomatic people transmit the virus for an average of $1/\delta + 1/\gamma$ days. We further assume that the virus can be detected by swab testing beyond the duration of the infectious period; this assumption is compatible with the hypothesis that transmission occurs for viral loads above a certain threshold but the diagnostic test can detect the presence of virus below the threshold for transmission. Compartments $\mathrm{TP_S}$ and $\mathrm{TP_A}$, respectively, represent symptomatic and asymptomatic people who are no longer infectious but have a detectable viral load, and hence test positive. Eventually, the viral load of all infections decreases below detection and people move into a test negative (TN) compartment. We assume a step change in the reproduction number on the day that lockdown started. Before the implementation of quarantine, the reproduction number is given by $R_0^1 = \beta\left(\frac{1}{\gamma} + \frac{1}{\delta}\right)$, and we assume that it drops to $R^2 = wR_0^1$ after the start of the lockdown, where $1 - w$ represents the per cent reduction in R_0^1 due to the intervention. We let T_i denote the number of participants swabbed on survey i ($i = 1, 2$) and let $P_{\mathrm{A}i}$, $P_{\mathrm{P}i}$ and $P_{\mathrm{s}i}$, respectively, denote the number of swabs testing positive among asymptomatic, presymptomatic (that is, those showing no symptoms at the time of testing but develop symptoms afterwards) and symptomatic participants, respectively. We assume that the number of positive swabs among symptomatic, presymptomatic and asymptomatic infections on survey i follows a binomial distribution with parameters T_i and π_{Xi}, where π_{Xi} represents the probability of testing positive on survey i for X (where $X = \mathrm{A, P, S}$). For symptomatic participants, $\pi_{\mathrm{S}i}$ is given by $\pi_{\mathrm{S}i} = \frac{I_{\mathrm{S}}(t_i) + \mathrm{TP_S}(t_i)}{N}$, for asymptomatic participants, $\pi_{\mathrm{A}i}$ is given by $\pi_{\mathrm{A}i} = \frac{p\mathrm{TP}(t_i) + I_{\mathrm{A}}(t_i) + \mathrm{TP_A}(t_i)}{N}$, and for presymptomatic participants, $\pi_{\mathrm{P}i}$ is given by $\pi_{\mathrm{P}i} = \frac{(1 - p)\mathrm{TP}(t_i)}{N}$, assuming perfect diagnostic sensitivity and specificity. The likelihood of the model is given by the product of the binomial distributions for symptomatic, presymptomatic and asymptomatic participants at times t_i, $i = 1, 2$. Inference was conducted in a Bayesian framework, using the Metropolis–Hastings Markov chain Monte Carlo (MCMC) method with uniform prior distributions[23]. We fixed the average generation time (equal to $1/\nu + 1/\delta + 1/\gamma$) to 7 days[19] and let the model infer $1/\nu$ and $1/\delta$. We explored the following values of R_0^1: 2.1, 2.4, 2.7, which are compatible with a doubling time of 3–4 days, as observed in Vo' and elsewhere in the Veneto region. We assumed that seeding of the infection occurred on 4 February 2020. We explored different scenarios on the average duration of viral detectability beyond the infectious period and fixed $1/\sigma$ to be 2, 4, 6, 8, 10 and 12 days. We estimate the number of infections introduced in the population from elsewhere at time t_0 (4 February 2020), the proportion of asymptomatic infections p, the average durations $1/\nu$, $1/\delta$ and $1/\gamma$ and the per cent reduction in R_0^1 due to the interventions $(1 - w)100\%$.

Analysis of associations

We applied logistic regression to test the association between SARS-CoV-2 positivity (overall and at the first and second surveys separately) with the age group (10 years of age bands, from 0 to >81 years of age) and sex (male and female). We used Fisher's exact test for comparing two binomial proportions to assess whether there is an association between the presence of symptoms for 41 confirmed COVID-19 cases who are resident in Vo' and the different types of comorbidities and treatments used. The analyses were repeated on the subset of patients who became negative at the second time point (results not shown). To increase the power of the data, we increased the sample size by including an additional 11 confirmed COVID-19 cases who were resident in other villages close to Vo'. None of these scenarios provided significant associations at the 5% level.

Ethical approval statement

The first sampling of the Vo' population was conducted within the surveillance programme established by the Veneto region and did not require ethical approval; the second sampling was approved by the Ethics Committee for Clinical Research of the province of Padua. Study participation was by consent. For participants under 18 years of age, consent was provided by a parent or legal guardian.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data set is available at https://github.com/ncov-ic/SEIR_Covid_Vo. Queries can be addressed to A.C. (a.drcrisanti@imperial.ac.uk; andrea.crisanti@unipd.it) or I.D. (i.dorigatti@imperial.ac.uk).

Code availability

The code is available at https://github.com/ncov-ic/SEIR_Covid_Vo.

18. Centers for Disease Control and Prevention. Interim guidelines for collecting, handling, and testing clinical specimens from persons for coronavirus disease 2019 (COVID-19). *CDC* https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html (accessed 18 May 2020).
19. Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. **25**, 2000045 (2020).
20. Python language reference, version 2.7 (Python Software Foundation, 2020).
21. R Core Team. R: A Language and Environment for Statistical Computing. http://www.R-project.org/ (R Foundation for Statistical Computing, 2020).
22. Nagraj, V. P. et al. epicontacts: handling, visualisation and analysis of epidemiological contacts. *F1000Res*. **7**, 566 (2018).
23. Robert, C. The Metropolis–Hastings algorithm. *Wiley StatsRef* https://doi.org/10.1002/9781118445112.stat07834 (2015).

Acknowledgements We thank the Mayor of the municipality of Vo', G. Martini, for his unreserved support throughout the study; a special thanks to the population of Vo' who volunteered en masse to this study; M. Perilli and S. Guglielmo for assistance in data collection and consistency check; and F. Bosa and G. Rupolo from the Italian Red Cross for the support in patient samplings. This work was supported by the Veneto region and was jointly funded by the UK Medical Research Council (MRC; grant MR/R015600/1), the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, the Abdul Latif Jameel Foundation and is also part of the EDCTP2 programme supported by the European Union. I.D. acknowledges research funding from a Sir Henry Dale Fellowship funded by the Royal Society and Wellcome Trust (grant 213494/Z/18/Z). C.C. acknowledges funding from the Wellcome Trust (grant 203851/Z/16/Z). L.C.O. from the Imperial College COVID-19 Response Team and G.C.-D. acknowledge research funding from The Royal Society. L.B., E.L. and S.T. acknowledge research funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement no. 874735 (VEO).

Author contributions A.C. conceived the project with input from E.L. and I.D. I.D. conceived the modelling with input from N.M.F. and C.A.D. E.L. coordinated data collection, curation and analyses. E.F. coordinated the diagnostic team and facilities. C.C. and G.C.-D. are joint second authors. E.F., L.B., C.D.V., L.R., R.M., A.L., D.A., M.S., E.D.C., M.C.V., F.S., F.O., V. Besutti, M.P., S.G.P., G.M. and M.T. performed laboratory testing on swabs and validated the results. E.L., S.T., V. Baldo, A.S., N.N. and S.C. analysed the data, contributed to the discussion and commented on the manuscript. A.R.B., I.D. and C.A.D. performed the statistical analyses. C.C., L.C., N.M.F. and I.D. developed the mathematical model. G.C.-D., K.A.M.G., C.A.D. and I.D. performed cluster analysis. E.L., M.N., F.C., G. Castelli, E.N., B.L., L. Fava and M.D. performed data collection, direct contacting of study participants at follow up and consistency check on metadata. S.M., R.S., G. Carretta, D.D. and L. Flor organized sampling logistics. S.M. and R.S. performed swab samplings. The Imperial College COVID-19 Response Team contributed to the discussion and background understanding of COVID-19 epidemiology. A.C. and I.D. wrote the manuscript, with contribution from E.L., L.B., V. Baldo and C.A.D.

Competing interests The authors declare no competing interests.

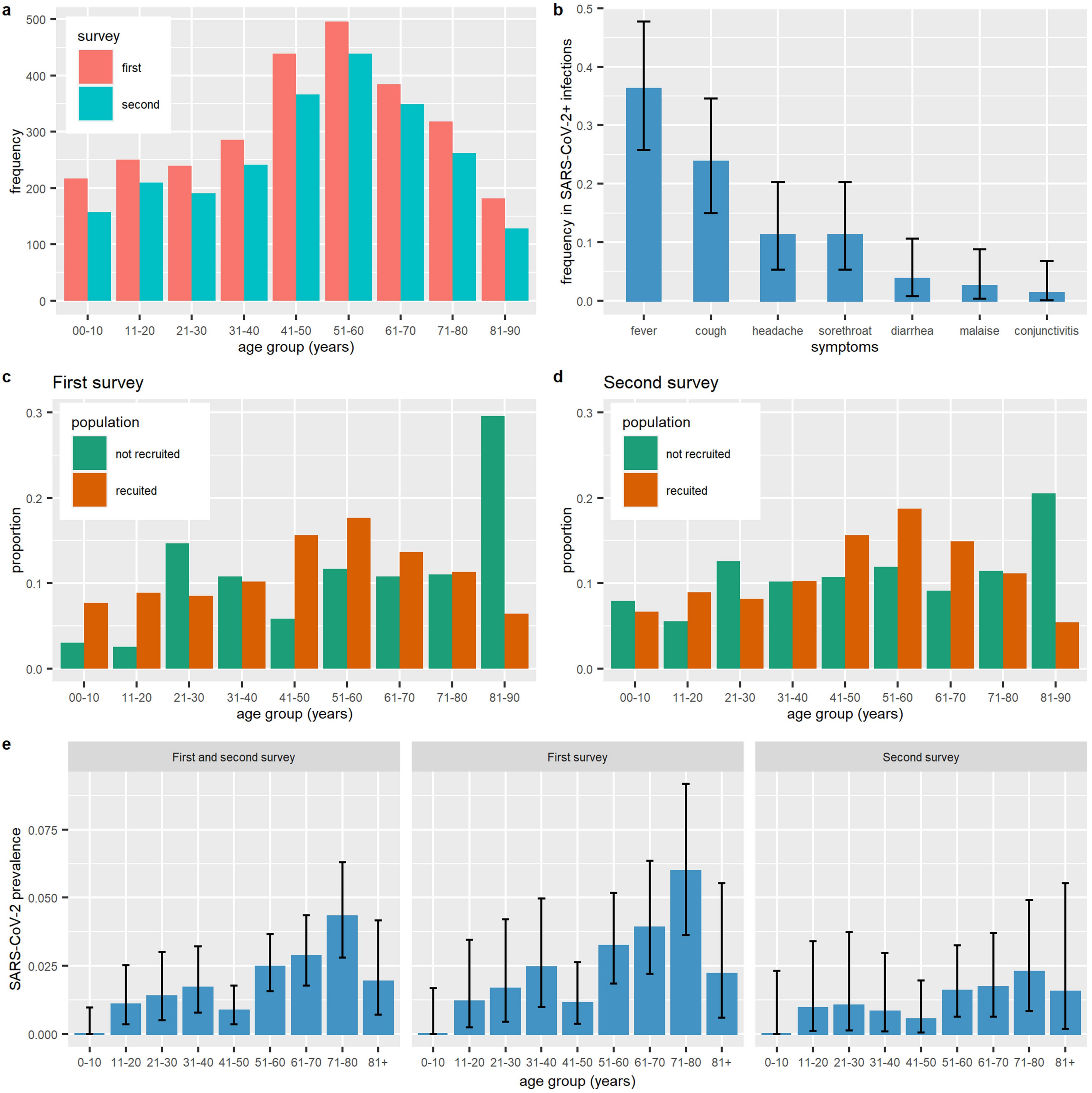
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2488-1.

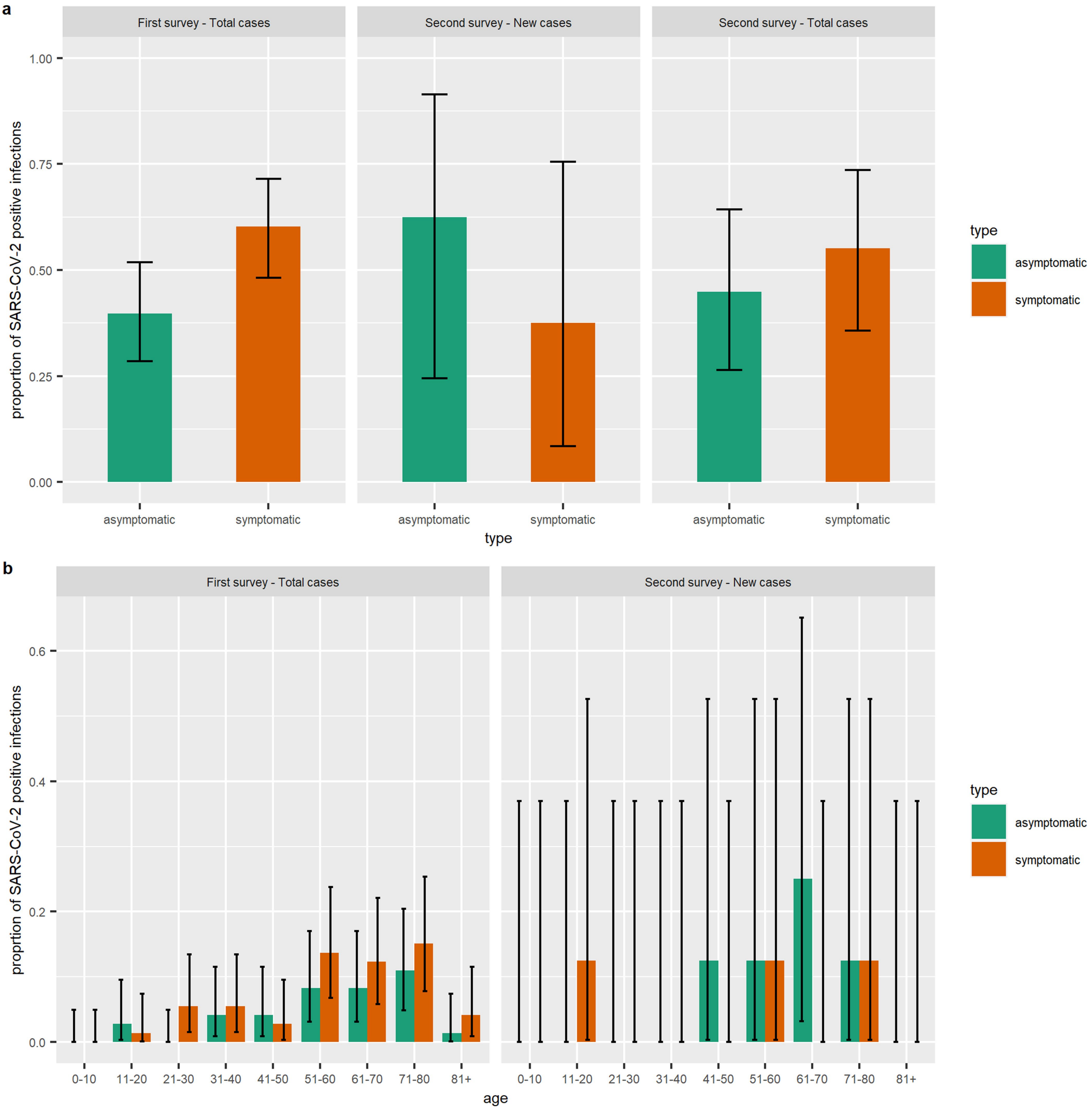
Correspondence and requests for materials should be addressed to I.D. or A.C.

Peer review information *Nature* thanks Gabriel Leung, Malik Peiris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

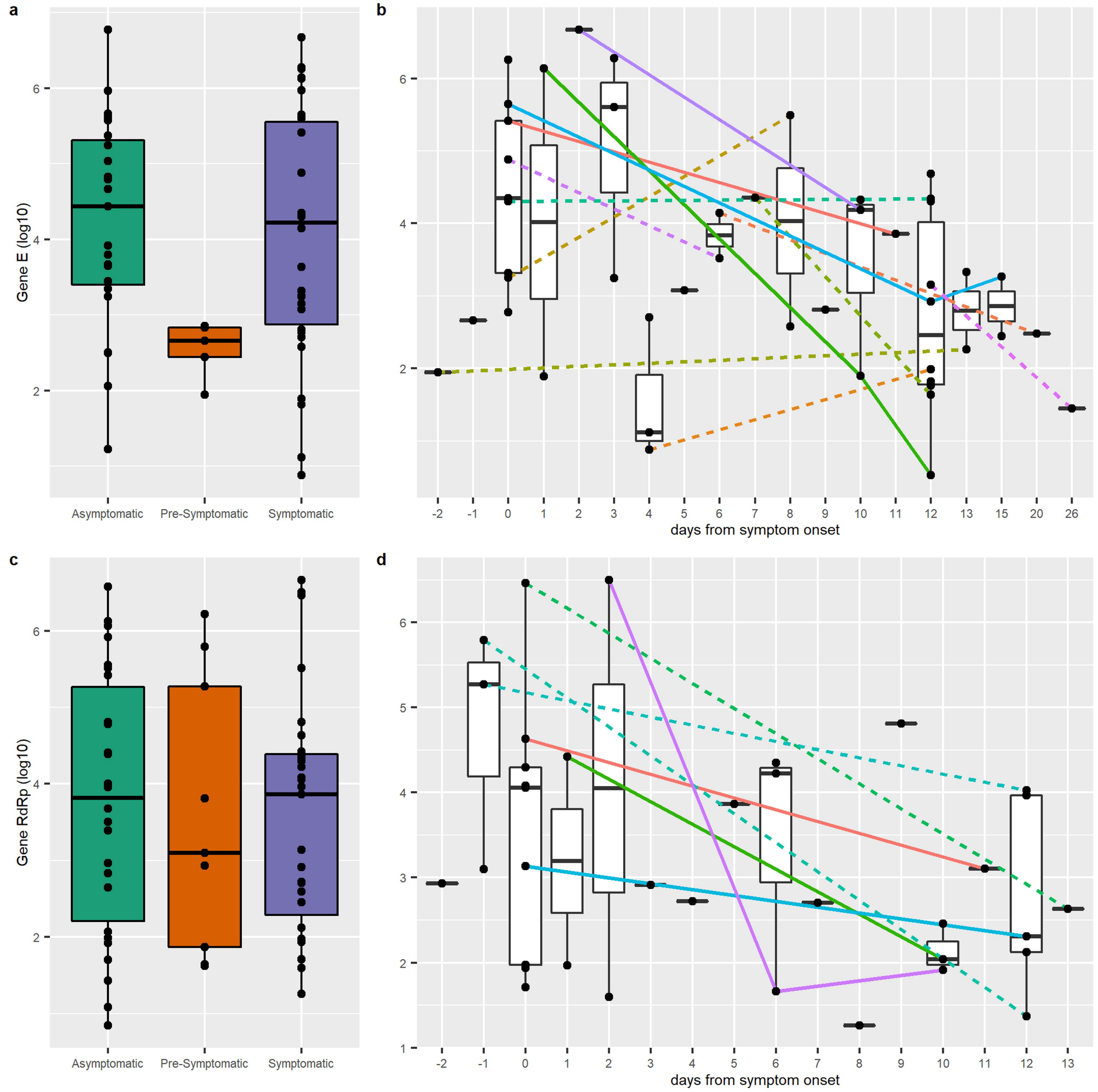
Reprints and permissions information is available at http://www.nature.com/reprints.



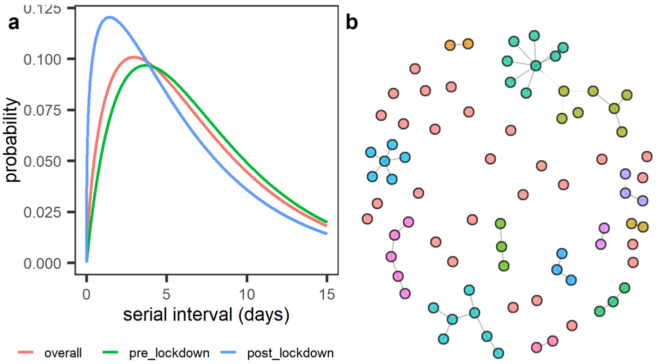
Extended Data Fig. 1 | Summary statistics, frequency of symptoms and prevalence by age. a, Age distributions (in years) of the participants enrolled in the first and second surveys. **b**, Frequency of individual symptoms (fever $x = 29$, cough $x = 19$, sore throat $x = 9$, headache $x = 9$, diarrhoea $x = 3$, malaise $x = 2$ and conjunctivitis $x = 1$) among participants with confirmed SARS-CoV-2 infection across the whole study period (that is, the first and second surveys aggregated; $n = 80$ participants). The error bars represent the 95% exact binomial CI. **c**, Age distribution of the population recruited and not recruited in the first survey. **d**, Age distribution of the population recruited and not recruited in the second survey. **e**, SARS-CoV-2 prevalence by age at the first and second surveys combined (positive $x = 0, 5, 6, 9, 7, 23, 21, 25$ and 6, tested $n = 374, 460, 431, 527, 805, 935, 733, 580$ and 310, respectively, in age groups 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81+ years) and at the first (positive $x = 0, 3, 4, 7, 5, 16, 15, 19$ and 4, tested $n = 217, 250, 240, 286, 439, 496, 384, 318$ and 182, respectively, in age groups 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81+ years) and second (positive $x = 0, 2, 2, 2, 2, 7, 6, 6$ and 2, tested $n = 157, 210, 191, 241, 366, 439, 389, 262$ and 128, respectively, in age groups 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81+ years) surveys separately. The error bars represent the 95% exact binomial CI.



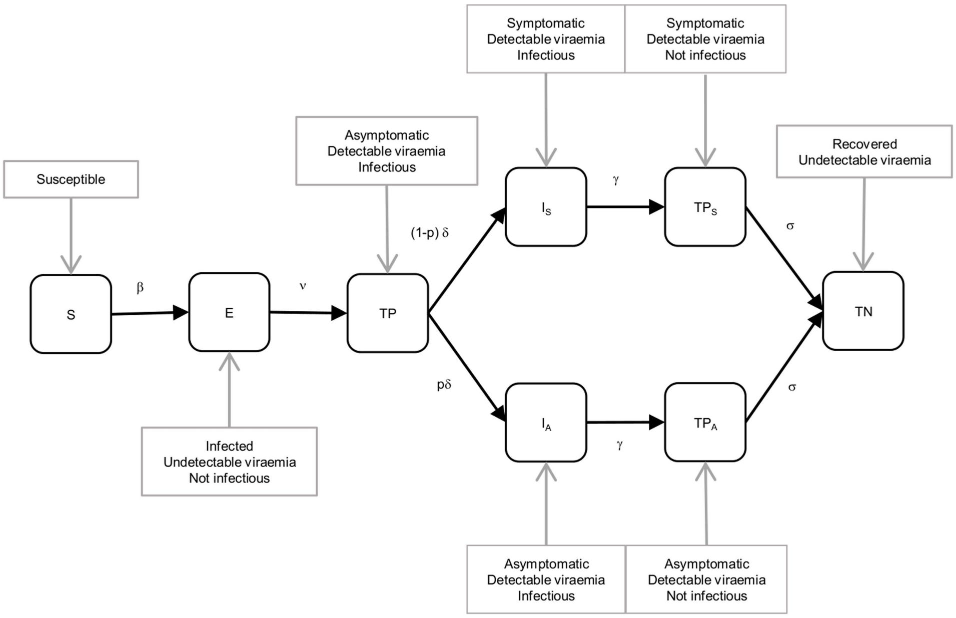
Extended Data Fig. 2 | Symptomatic and asymptomatic infection statistics. **a**, Relative proportion of asymptomatic and symptomatic SARS-CoV-2 infections among the total number of positive swabs in the first survey (first survey – total cases; asymptomatic $x = 29$, symptomatic $x = 44$, tested $n = 73$), second survey (second survey – total cases; asymptomatic $x = 13$, symptomatic $x = 16$, tested $n = 29$) and among the number of new positive swabs in the second survey (second survey – new cases; asymptomatic $x = 5$, symptomatic $x = 3$, tested $n = 8$). The error bars represent the 95% exact binomial CI. **b**, Age distribution and relative proportion of asymptomatic and symptomatic SARS-CoV-2-positive infections among the total number of positive swabs in the first survey (first survey – total cases; asymptomatic $x = 0, 2, 0, 3, 3, 6, 6, 8$ and 1, symptomatic $x = 0, 1, 4, 4, 2, 10, 9, 11$ and 3, respectively, in age groups 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81+ years; tested $n = 73$) and among the number of new positive swabs in the second survey (second survey – new cases; asymptomatic $x = 0, 0, 0, 0, 1, 1, 2, 1$ and 0, symptomatic $x = 0, 1, 0, 0, 0, 1, 0, 1$ and 0, respectively, in age groups 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81+ years; tested $n = 8$). The error bars represent the 95% exact binomial CI.



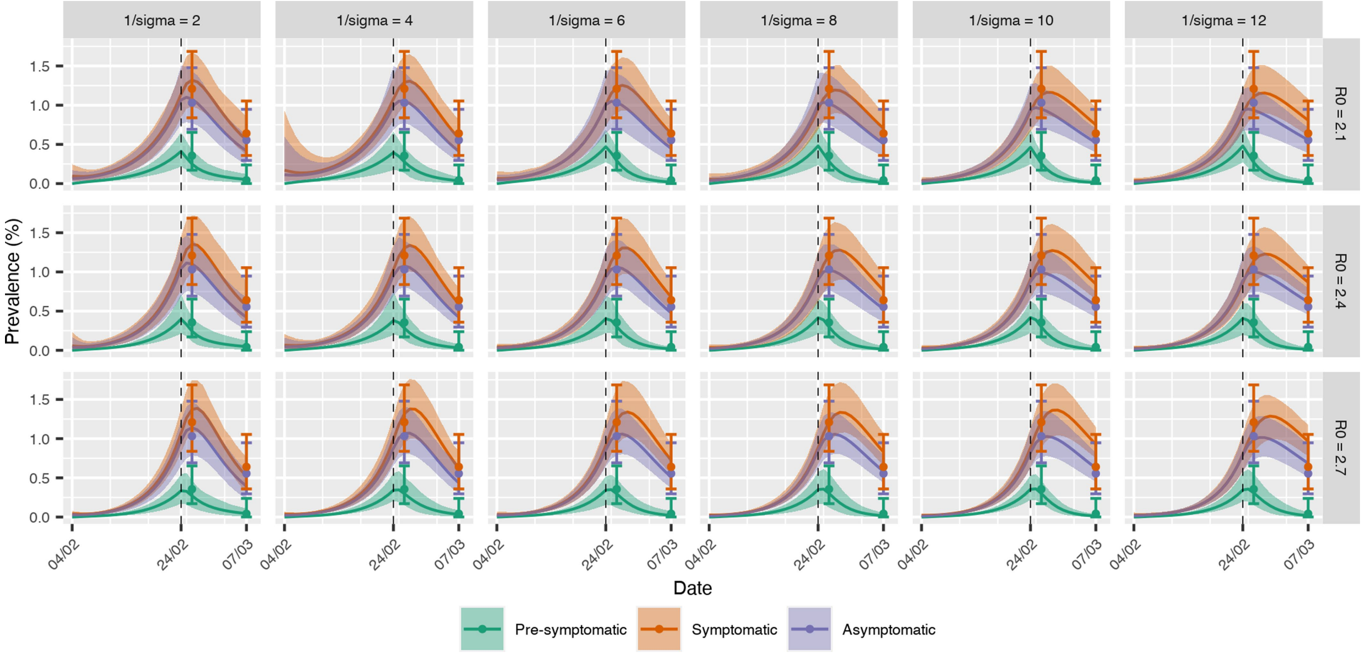
Extended Data Fig. 3 | Viral load for asymptomatic, pre-symptomatic and symptomatic infections and viral load dynamics relative to the number of days from symptom onset. **a**, The median (solid line), the interquartile range (that is, 25th to 75th percentiles (box)) and the range (that is, minimum to maximum (whiskers)) of gene *E* genome equivalent copies per ml ($\log_{10}$ scale, y axis) calculated from RT–PCR interpolated values (asymptomatic $n = 23$, pre-symptomatic $n = 5$ and symptomatic $n = 30$). The raw C_t data and the derived values of the genome equivalent copies are provided in the data set. **b**, The median (solid line), the interquartile range (that is, 25th to 75th percentiles (box)) and the range (that is, minimum to maximum (whiskers)) of gene *E* genome equivalent copies per ml ($\log_{10}$ scale, y axis) versus the number of days from symptom onset (days, x axis); $n = 34$ participants. The lines in colour join measurements from the same participant. The solid lines identify the four participants with sequential viral load measurements for both gene *E* and gene *RdRp*. **c**, The median (solid line), the interquartile range (that is, 25th to 75th percentiles (box)) and the range (that is, minimum to maximum (whiskers)) of *RdRp* genome equivalent copies per ml ($\log_{10}$ scale, y axis) calculated from RT–PCR interpolated values (asymptomatic $n = 26$, pre-symptomatic $n = 9$ and symptomatic $n = 27$). The raw C_t data and the derived values of genome equivalent copies are provided in the data set. **d**, The median (solid line), the interquartile range (that is, 25th to 75th percentiles (box)) and the range (that is, minimum to maximum (whiskers)) of *RdRp* genome equivalent copies per ml ($\log_{10}$ scale, y axis) versus the number of days from symptom onset (days, x axis); $n = 28$ participants. The lines in colour join measurements from the same participant. The solid lines identify the four participants with sequential viral load measurements for both gene *E* and gene *RdRp*.



Extended Data Fig. 4 | Serial interval distribution and transmission chains. **a**, Estimated serial interval distributions for the whole study period (overall) and for the pre-lockdown (before 24 February 2020) and post-lockdown (after 24 February 2020) periods. **b**, Observed transmission clusters from reported and household contacts. Each node (circle) represents a positive infection, and the edges (the line connecting the nodes) connect positive infections that reported contacts or are household members. The different colours represent different clusters of infection.



Extended Data Fig. 5 | Flow chart of the mathematical model fitted to the point prevalence data observed in Vo' at the first and second surveys. Further details are provided in the Methods.



Extended Data Fig. 6 | SARS-CoV-2 dynamics in Vo' inferred from the fit of the dynamical model to the observed prevalence of symptomatic, pre-symptomatic and asymptomatic infections in the first and second surveys. Each sub-panel represents the model fit using the specified values of R_0^1 (the reproduction number before the lockdown) and $1/\sigma$ (the average duration of positivity beyond the duration of the infectious period). The dashed vertical line represents the time that lockdown started. The points represent the observed prevalence data; the 95% CI is the exact binomial CI. The solid lines represent the mean and the shading represents the 95% credible interval obtained from 100 samples from the posterior distribution of the parameters.

Extended Data Table 1 | Age distribution of Vo' residents and the number of tested participants at the two time points across different age groups

Age group (years)	Resident subjects	First survey		Second survey	
		n.	(%)	n.	(%)
00-10	231	217	93,9	157	68,0
11-20	262	250	95,4	210	80,2
21-30	308	240	77,9	191	62,0
31-40	336	286	85,1	241	71,7
41-50	466	439	94,2	366	78,5
51-60	550	496	90,2	439	79,8
61-70	434	384	88,5	349	80,4
71-80	369	318	86,2	262	71,0
81+	319	182	57,1	128	40,1
total	3275	2812	85,9	2343	71,5

Extended Data Table 2 | Age distribution of symptomatic and asymptomatic individuals at the first and second surveys

Age group	Tested at first survey		Positive at first survey				Tested at second survey		Positive at second survey							
									Total cases				New cases only			
	Symp	Asymp	Symp^	(%)	Asymp^	(%)	Symp	Asymp	Symp	(%)	Asymp	(%)	Symp	(%)	Asymp	(%)
00-10	28	189	-	(-)	-	(-)	15	142	-	(-)	-	(-)	-	(-)	-	(-)
11-20	24	226	1	(4.2)	2 (1)	(0.9)	22	188	2	(9.1)	-	(-)	1	(4.5)	-	(-)
21-30	14	226	4 (2)	(28.6)	0	(-)	10	181	2	(20.0)	-	(-)	-	(-)	-	(-)
31-40	23	263	4	(17.4)	3	(1.1)	20	221	-	(-)	2	(0.9)	-	(-)	-	(-)
41-50	27	412	2	(7.4)	3 (1)	(0.7)	27	339	-	(-)	2	(0.6)	-	(-)	1	(0.3)
51-60	32	464	10	(31.3)	6 (1)	(1.3)	28	411	5	(17.9)	2	(0.5)	1	(3.6)	1	(0.2)
61-70	16	368	9	(56.3)	6	(1.6)	16	333	2	(12.5)	4	(1.2)	-	(-)	2	(0.6)
71-80	21	297	11 (1)	(52.4)	8 (1)	(2.7)	15	247	3	(20.0)	3	(1.2)	1	(6.7)	1	(0.4)
81+	8	174	3	(37.5)	1 (1)	(0.6)	8	120	2	(25.0)	-	(-)	-	(-)	-	(-)
Total	193	2619	44	(22.8)	29	(1.1)	161	2182	16	(9.9)	13	(0.6)	3	(1.9)	5	(0.2)

The symptomatic (Symp) category includes both symptomatic and pre-symptomatic participants. The percentages represent the proportions positive among those tested; that is, the probability of testing positive given symptomatic or asymptomatic (Asymp) infection.

^Participants not available at the second survey are reported within parentheses.

Extended Data Table 3 | Children negative for SARS-CoV-2 living in households with infected relatives

	first survey	second survey
n (age group 0-10)	217	157
with positive cohabitant*	10	3
with positive relative not cohabitant§	2	0

*Five participants are residents outside Vo' and are not included in the released data set.

§Both participants did not reside in Vo' and were not included in the released data set.

Extended Data Table 4 | Results of the second survey for participants living with or reporting close contacts with relatives infected with SARS-CoV-2

		Second survey			
		New cases		Negative	
		n.	(%)	n.	(%)
Subjects living with or reporting close contacts with infected relatives	Yes	6	(75.0)	78	(3.4)
	No	2	(25.0)	2197	(96.6)
Total		8		2275	

nature research

Corresponding author(s): Ilaria Dorigatti and Andrea Crisanti

Last updated by author(s): Jun 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collated using Microsoft Excel (version 16.16.6).
Data analysis	We developed code in the R programming language (version 3.6.0) using packages tidyr, dplyr, readr, janitor, lubridate, magrittr, ggplot2, plotly, htmlwidgets, epicontacts, igraph, GGally, epitrix, ggpubr, flextable, network, sna, scales, intergraph, cowplot, readxl, odin, stringr, ggrepel, gridExtra, sf, ggspatial and in Python (version 3.7.3) using modules scipy 1.4.1, numpy 1.18.1, and matplotlib 3.2.1. The code is available at https://github.com/ncov-ic/SEIR_Covid_Vo.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The dataset is available at https://github.com/ncov-ic/SEIR_Covid_Vo.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Vo' has a resident population of 3,275 inhabitants. We collected nasopharyngeal swabs from 2,812 and 2,343 subjects in the first and second screening respectively, corresponding to 85.9% and 71.5% of the eligible population. No sample size calculation was performed, we aimed to recruit as many residents as possible.
Data exclusions	We excluded from the analysis the data collected on a small number of subjects, including 11 confirmed COVID-19 infections, who did not reside in Vo'.
Replication	Detection of SARS-CoV-2 RNA was performed by an in-house real-time RT-PCR method performed at the Clinical Microbiology and Virology Unit of Padova University Hospital, which is the Regional Reference Laboratory for emerging viral infections. The samples collected in the initial phase of the survey were validated by the National Reference Laboratory at the Italian Institute of Health (Istituto Superiore di Sanità) and demonstrated 100% agreement with the in-house assay. Given the 100% agreement on the samples collected in the initial phase and due to the large number of samples analyzed by the laboratory during the epidemic, we did not validate all samples collected in Vo' across the two surveys.
Randomization	Randomization is not relevant in our study, we aimed to enroll as many study participants as possible.
Blinding	Blinding is not relevant in our study, it was an observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	We collected information on sampling dates, results of SARS-CoV-2 testing, age, sex, symptoms, underlying health conditions, pharmacological therapy, hospitalization, household composition and contact network. The recruited subjects were between 1 month and 100 years of age and 49.9% were male and 50.1% were female. The underlying health conditions and pharmacological therapies of the recruited population at the time of the study are described in Supplementary Tables S3 and S4.
Recruitment	Study participation was by consent. For subjects under the age of 18 years, consent was provided by a parent or legal guardian. Participation in the study was publicized through local authorities. The age distribution of the recruited versus not recruited population was statistically different, as described in the main text and in Extended Data Figure 1.
Ethics oversight	The Ethics Committee for Clinical Research of the province of Padova approved the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Factors associated with COVID-19-related death using OpenSAFELY

https://doi.org/10.1038/s41586-020-2521-4

Received: 15 May 2020

Accepted: 1 July 2020

Elizabeth J. Williamson[1,6], Alex J. Walker[2,6], Krishnan Bhaskaran[1,6], Seb Bacon[2,6], Chris Bates[3,6], Caroline E. Morton[2], Helen J. Curtis[2], Amir Mehrkar[2], David Evans[2], Peter Inglesby[2], Jonathan Cockburn[3], Helen I. McDonald[1,4], Brian MacKenna[2], Laurie Tomlinson[1], Ian J. Douglas[1], Christopher T. Rentsch[1], Rohini Mathur[1], Angel Y. S. Wong[1], Richard Grieve[1], David Harrison[5], Harriet Forbes[1], Anna Schultze[1], Richard Croker[2], John Parry[3], Frank Hester[3], Sam Harper[3], Rafael Perera[2], Stephen J. W. Evans[1], Liam Smeeth[1,4,7] & Ben Goldacre[2,7] ✉

Coronavirus disease 2019 (COVID-19) has rapidly affected mortality worldwide[1]. There is unprecedented urgency to understand who is most at risk of severe outcomes, and this requires new approaches for the timely analysis of large datasets. Working on behalf of NHS England, we created OpenSAFELY—a secure health analytics platform that covers 40% of all patients in England and holds patient data within the existing data centre of a major vendor of primary care electronic health records. Here we used OpenSAFELY to examine factors associated with COVID-19-related death. Primary care records of 17,278,392 adults were pseudonymously linked to 10,926 COVID-19-related deaths. COVID-19-related death was associated with: being male (hazard ratio (HR) 1.59 (95% confidence interval 1.53–1.65)); greater age and deprivation (both with a strong gradient); diabetes; severe asthma; and various other medical conditions. Compared with people of white ethnicity, Black and South Asian people were at higher risk, even after adjustment for other factors (HR 1.48 (1.29–1.69) and 1.45 (1.32–1.58), respectively). We have quantified a range of clinical factors associated with COVID-19-related death in one of the largest cohort studies on this topic so far. More patient records are rapidly being added to OpenSAFELY, we will update and extend our results regularly.

On 11 March 2020, the World Health Organization (WHO) characterized COVID-19—which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—as a pandemic, after 118,000 cases and 4,291 deaths were reported in 114 countries[2]. As of 6 May 2020 (the date of latest data availability for this study), cases had reached over 3.5 million globally, with more than 240,000 deaths attributed to the virus[1]. On the same day in the UK, there had been 206,715 confirmed cases of COVID-19, and 30,615 COVID-19-related deaths[3].

Age and gender are well-established risk factors for severe COVID-19 outcomes: over 90% of the COVID-19-related deaths in the UK have been in people over 60, and 60% in men[4]. Various pre-existing conditions have also been associated with increased risk. For example, the Chinese Center for Disease Control and Prevention reported in a study of 44,672 individuals (1,023 deaths) that cardiovascular disease, hypertension, diabetes, respiratory disease and cancers were associated with an increased risk of death[5]; however, correction for relationships with age was not possible. A UK cross-sectional survey of 16,749 patients who were hospitalized with COVID-19 showed that the risk of death was higher for patients with cardiac, pulmonary and kidney disease, as well as cancer, dementia and obesity (HRs of 1.19–1.39 after correction for age and sex)[6]. Obesity was associated with treatment escalation in a French intensive care cohort[7] ($n = 124$) and a New York hospital presentation cohort[8] ($n = 3{,}615$). The risks associated with smoking are unclear[9–11]. People from Black and minority ethnic groups are at increased risk of poor outcomes from COVID-19, for reasons that are unclear[12,13].

Patient care is typically managed through electronic health records, which are commonly used in research. However traditional approaches to the analysis of electronic health records rely on intermittent extracts of small samples of historic data. Evaluating a rapidly arising novel cause of death requires a new approach. We therefore set out to deliver a secure analytics platform inside the data centre of major electronic health records vendors, running across the full, linked and pseudonymized electronic health records of a very large population of NHS patients, to determine factors that are associated with COVID-19-related death in England.

Associations with COVID-19-related death

In total, 17,278,392 adults were included (Fig. 1; cohort description in Table 1). Eleven per cent of individuals (1,851,868) had ethnicity recorded as mixed, South Asian, Black or other (hereafter referred to

---

[1]London School of Hygiene and Tropical Medicine, Faculty of Epidemiology and Population Health, London, UK. [2]The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. [3]TPP, Horsforth, UK. [4]NIHR Health Protection Research Unit in Immunisation, London, UK. [5]Intensive Care National Audit and Research Centre (ICNARC), London, UK. [6]These authors contributed equally: Elizabeth J. Williamson, Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates. [7]These authors jointly supervised this work: Liam Smeeth, Ben Goldacre. ✉e-mail: ben.goldacre@phc.ox.ac.uk

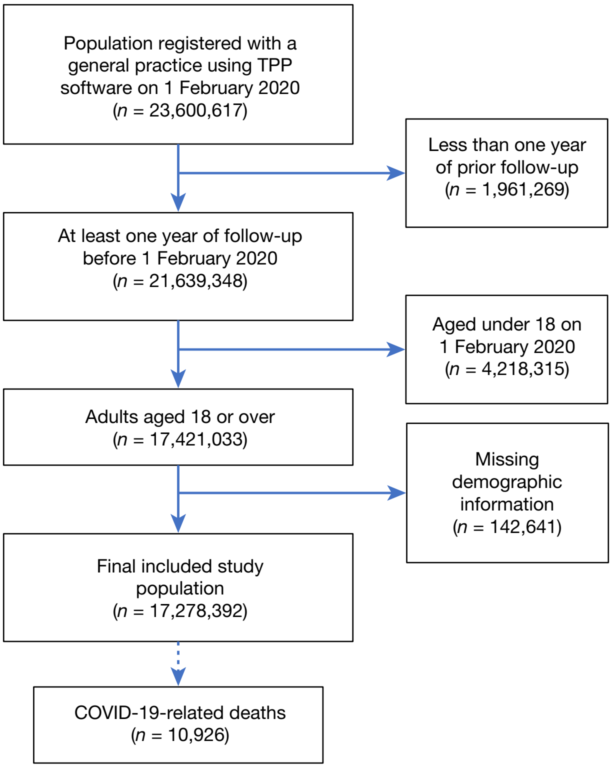


Fig. 1 | Flow diagram of the cohort. The diagram shows the numbers of individuals (*n*) excluded at different stages and the identification of cases for the main end points.

as Black and minority ethnic, BAME). There were missing data for body mass index (3,751,769; 22%), smoking status (720,923; 4%), ethnicity (4,560,113; 26%) and blood pressure (1,715,095; 10%). COVID-19-related death was recorded in linked death registration data for 10,926 of the study population.

The overall cumulative incidence of COVID-19-related death 90 days after the start of the study was less than 0.01% in those aged 18–39 years, rising to 0.67% and 0.44% in men and women, respectively, aged 80 years or over (Fig. 2).

Associations between patient-level factors and risk of COVID-19-related death are shown in Table 2 and Fig. 3. Increasing age was strongly associated with risk, with people aged 80 or over having a more than 20-fold-increased risk compared to 50–59-year-olds (fully adjusted HR 20.60; 95% confidence interval (CI) 18.70–22.68). With age fitted as a flexible spline, an approximately log-linear relationship was observed (Extended Data Fig. 1). Men had a higher risk than women (fully adjusted HR 1.59 (1.53–1.65)). These findings are consistent with patterns observed in smaller studies worldwide and in the UK[14].

People from all BAME groups were at higher risk than those of white ethnicity. When adjusted only for age and sex, hazard ratios ranged from 1.62–1.88 for Black and South Asian individuals and people of mixed ethnicities, compared to white people, decreasing to 1.43–1.48 after adjustment for all included factors (results for more detailed categories are shown in Extended Data Table 1). BAME ethnicity has previously been found to be associated with an increased risk of COVID-19 infection and poor outcomes[12,13,15]. Our findings show that only a small part of the excess risk is explained by a higher prevalence of medical problems such as cardiovascular disease or diabetes among BAME people, or by higher levels of deprivation.

We found a consistent pattern of increasing risk with greater deprivation, with the most deprived quintile having a hazard ratio of 1.79 compared to the least deprived, consistent with recent national statistics[16]. Again, very little of this increased risk was explained by pre-existing disease or clinical factors, suggesting that other social factors have an important role.

Increasing risks were seen with increasing obesity (fully adjusted HR 1.92 (1.72–2.13) for a body mass index (BMI; $\mathrm{kg\,m^{-2}}$) of over 40), and most comorbidities were associated with a higher risk of COVID-19-related death, including diabetes (greater hazard ratio for those with a recently measured glycated haemoglobin (HbA1c) level of at least 58 $\mathrm{mmol\,mol^{-1}}$), severe asthma (defined as asthma with recent use of an oral corticosteroid), respiratory disease, chronic heart disease, liver disease, stroke, dementia, other neurological diseases, reduced kidney function (greater hazard ratio associated with a lower estimated glomerular filtration rate; eGFR), autoimmune diseases (rheumatoid arthritis, lupus or psoriasis) and other immunosuppressive conditions (Table 2). Those with a recent (that is, in the last five years) history of haematological malignancy had an at least 2.5-fold increased risk, which decreased slightly after five years. For other cancers, hazard ratios were smaller and increased risks were associated mainly with recent diagnoses. A history of dialysis or end-stage renal failure was associated with increased risk when added in a secondary analysis (HR 3.69 (3.09–4.39)). These findings largely concur with other data, including the UK international severe acute respiratory and emerging infection consortium (ISARIC) study of hospitalized UK patients with COVID-19—which indicated an increased risk of death with cardiac, pulmonary and kidney disease, malignancy, obesity and dementia[6]—and a large Chinese study that, although lacking correction for age, suggested that cardiovascular disease, hypertension, diabetes, respiratory disease and cancers are associated with increased mortality[5]. Our results showing that severe asthma is associated with a higher risk are notable, as early data suggested that asthma was under-represented in patients with COVID-19 who were hospitalized or had severe outcomes[17,18].

Post hoc analyses of smoking and hypertension

Both current and former smoking were associated with a higher risk in models that were adjusted for age and sex only, but in the fully adjusted model current smoking was associated with a lower risk (fully adjusted HR 0.89 (0.82–0.97)), which concurs with the lower than expected prevalence of smoking that was observed in previous studies among patients with COVID-19 in China[10], France[11] and the United States[19]. We investigated this in more depth post hoc by adding covariates individually to the age, sex and smoking model, and found that the change in hazard ratio was driven largely by adjustment for chronic respiratory disease (HR 0.98 (0.90–1.06) after adjustment). This and other comorbidities could be consequences of smoking, highlighting that the fully adjusted smoking hazard ratio cannot be interpreted causally owing to the inclusion of factors that are likely to mediate smoking effects. We therefore then fitted a model adjusted for demographic factors only (age, sex, deprivation and ethnicity), which showed a non-significant positive hazard ratio for current smoking (HR 1.07 (0.98–1.18)). This does not support any postulated protective effect of nicotine[9,20], but suggests that any increased risk with current smoking is likely to be small and will need to be clarified as the epidemic progresses and more data accumulate.

We similarly investigated the change in the hypertension hazard ratio (from 1.09 (1.05–1.14) adjusted for age and sex, to 0.89 (0.85–0.93) with all covariates included), and found that diabetes and obesity were principally responsible for this reduction (HR 0.97 (0.92–1.01) adjusted for age, sex, diabetes and obesity). Given the strong association between blood pressure and age we then examined the interaction between these variables; this revealed strong evidence of interaction ($P < 0.001$), with hypertension associated with a higher risk up to the age of 70 years and a lower risk above the age of 70 (adjusted HRs 3.10 (1.69–5.70), 2.73 (1.96–3.81), 2.07 (1.73–2.47), 1.32 (1.17–1.50), 0.94 (0.86–1.02) and 0.73 (0.69–0.78) for ages 18–39, 40–49, 50–59, 60–69, 70–79 and 80 or over, respectively). The reasons for the inverse association

Table 1 | Cohort description with number of COVID-19 deaths by patient characteristics

Characteristic	Category	Number of individuals (column %)	Number of COVID-19-related deaths (% within stratum)
Total		17,278,392 (100.0)	10,926 (0.06)
Age	18–39	5,914,384 (34.2)	54 (0.00)
	40–49	2,849,984 (16.5)	140 (0.00)
	50–59	3,051,110 (17.7)	522 (0.02)
	60–69	2,392,392 (13.8)	1,101 (0.05)
	70–79	1,938,842 (11.2)	2,635 (0.14)
	80+	1,131,680 (6.5)	6,474 (0.57)
Sex	Female	8,647,989 (50.1)	4,764 (0.06)
	Male	8,630,403 (49.9)	6,162 (0.07)
BMI (kg m^{-2})	<18.5	310,721 (1.8)	522 (0.17)
	18.5–24.9	4,763,150 (27.6)	3,364 (0.07)
	25–29.9	4,682,906 (27.1)	3,068 (0.07)
	30–34.9 (obese class I)	2,384,406 (13.8)	1,813 (0.08)
	35–39.9 (obese class II)	922,398 (5.3)	762 (0.08)
	≥40 (obese class III)	463,042 (2.7)	379 (0.08)
	Missing	3,751,769 (21.7)	1,018 (0.03)
Smoking	Never	7,924,739 (45.9)	3,598 (0.05)
	Former	5,690,966 (32.9)	6,531 (0.11)
	Current	2,941,764 (17.0)	708 (0.02)
	Missing	720,923 (4.2)	89 (0.01)
Ethnicity	White	10,866,411 (62.9)	7,119 (0.07)
	Mixed	169,697 (1.0)	62 (0.04)
	South Asian	1,022,130 (5.9)	608 (0.06)
	Black	339,909 (2.0)	250 (0.07)
	Other	320,132 (1.9)	110 (0.03)
	Missing	4,560,113 (26.4)	2,777 (0.06)
IMD quintile	1 (least deprived)	3,497,154 (20.2)	1,908 (0.05)
	2	3,476,668 (20.1)	2,030 (0.06)
	3	3,483,668 (20.2)	2,114 (0.06)
	4	3,480,459 (20.1)	2,388 (0.07)
	5 (most deprived)	3,340,443 (19.3)	2,486 (0.07)
Blood pressure	Normal	3,804,148 (22.0)	2,487 (0.07)
	Elevated	2,482,710 (14.4)	1,899 (0.08)
	High stage 1	5,548,198 (32.1)	3,281 (0.06)
	High stage 2	3,728,241 (21.6)	3,229 (0.09)
	Missing	1,715,095 (9.9)	30 (0.00)
High blood pressure or diagnosed hypertension		5,925,492 (34.3)	8,049 (0.14)
Respiratory disease excluding asthma		703,917 (4.1)	2,240 (0.32)
Asthma[a]	With no recent OCS use	2,454,403 (14.2)	1,211 (0.05)
	With recent OCS use	291,670 (1.7)	335 (0.11)
Chronic heart disease		1,167,455 (6.8)	3,811 (0.33)
Diabetes[b]	With HbA1c < 58 mmol mol^{-1}	1,038,082 (6.0)	2,391 (0.23)
	With HbA1c ≥ 58 mmol mol^{-1}	486,491 (2.8)	1,254 (0.26)
	With no recent HbA1c measure	193,993 (1.1)	444 (0.23)
Cancer (non-haematological)	Diagnosed <1 year ago	79,964 (0.5)	220 (0.28)
	Diagnosed 1–4.9 years ago	234,186 (1.4)	449 (0.19)
	Diagnosed ≥5 years ago	542,320 (3.1)	1,125 (0.21)
Haematological malignancy	Diagnosed <1 year ago	8,704 (0.1)	43 (0.49)
	Diagnosed 1–4.9 years ago	27,742 (0.2)	120 (0.43)
	Diagnosed ≥5 years ago	63,460 (0.4)	173 (0.27)

Continued

Characteristic	Category	Number of individuals (column %)	Number of COVID-19-related deaths (% within stratum)
Reduced kidney function[c]	eGFR 30–60	1,007,383 (5.8)	3,987 (0.40)
	eGFR < 30	78,093 (0.5)	864 (1.11)
Kidney dialysis		23,978 (0.1)	192 (0.80)
Liver disease		100,017 (0.6)	181 (0.18)
Stroke or dementia		390,002 (2.3)	2,423 (0.62)
Other neurological disease		170,448 (1.0)	665 (0.39)
Organ transplant		20,001 (0.1)	69 (0.34)
Asplenia		27,917 (0.2)	40 (0.14)
Rheumatoid arthritis, lupus or psoriasis		878,475 (5.1)	962 (0.11)
Other immunosuppressive condition		44,504 (0.3)	52 (0.12)

IMD, index of multiple deprivation.

[a] For oral corticosteroid (OCS) use, 'recent' refers to <1 year before baseline.

[b] Classification by HbA1c is based on measurements within 15 months of baseline.

[c] eGFR is measured in ml min^{-1} per 1.73 m^2 and taken from the most recent serum creatinine measurement.

between hypertension and mortality in older individuals are unclear and warrant further investigation, including detailed examination of frailty, comorbidity and drug exposures in this age group.

Model checking and sensitivity analyses

The average C-statistic—a measure of the model's ability to distinguish between patients who experience COVID-19-related deaths and those who do not, ranging from 0 (no ability) to 1 (perfect ability)—was 0.93. Results were similar when missing data were handled using analysis of complete records only, or using multiple imputation (sensitivity analyses; Extended Data Table 2). Non-proportional hazards were detected in the primary model ($P < 0.001$). A sensitivity analysis with earlier administrative censoring at 6 April 2020—before which mortality should not have been affected by the social distancing policies that were introduced in the UK in late March—showed no evidence of non-proportional hazards ($P = 0.83$). Hazard ratios were similar but somewhat larger in magnitude for some covariates, whereas the association with increasing deprivation appeared to be smaller (Extended Data Table 2).

Discussion

This secure analytics platform operating across NHS patient records of over 17 million adults and 6 million children was used to identify, quantify and analyse factors associated with COVID-19-related death in one of the largest cohort studies on this topic conducted by any country so far. Most comorbidities were associated with increased risk, including cardiovascular disease, diabetes, respiratory disease (including severe asthma), obesity, a history of haematological malignancy or recent other cancer, kidney, liver and neurological diseases, and autoimmune conditions. South Asian and Black people had a substantially higher risk of COVID-19-related death than white people, and this was only partly attributable to comorbidities, deprivation or other factors. A strong association between deprivation and risk was also only partly explained by comorbidities or other factors.

Our analyses provide a preliminary picture of how key demographic characteristics and a range of comorbidities—which were a priori selected as being of interest in COVID-19—are jointly associated with poor outcomes. These initial results may be used to inform the development of prognostic models. We caution against interpreting our estimates as causal effects. For example, the fully adjusted smoking hazard ratio does not capture the causal effect of smoking, owing to the inclusion of comorbidities that are likely to mediate any effect of smoking on COVID-19-related death (for example, chronic obstructive pulmonary disease). Our study has highlighted a need for carefully designed analyses that specifically focus on the causal effect of smoking on COVID-19-related death. Similarly, there is a need for analyses exploring the causal relationships that underlie the associations observed between hypertension and COVID-19-related death.

Strengths and weaknesses

The greatest strengths of this study are its size and the speed at which it was conducted. By building a secure analytics platform across routinely collected live clinical data stored in situ, we have produced timely results from the current NHS records of approximately 40% of the English population. The large scale of the study allows more precision—on rarer exposures and on multiple factors—and rapid detection of important signals. Our platform will expand to provide updated analyses over time. Another strength is our use of open methods: we pre-specified our analysis plan and shared our full analytic code and codelists for review and reuse. We ascertained patient demographics, medications and comorbidities from full pseudonymized longitudinal primary care records, which provide substantially more detail than data that are recorded on admission to hospital, and which take into account the total population rather than the selected subset of individuals who present at hospitals. We censored deaths from other causes using data from the UK Office for National Statistics (ONS). Analyses were stratified by area to account for known geographical differences in the incidence of COVID-19.

The study also has some important limitations. In our outcome definition, we included clinically suspected (non-laboratory-confirmed) cases of COVID-19, because testing has not always been carried out, especially in older patients in care homes. However, this may have

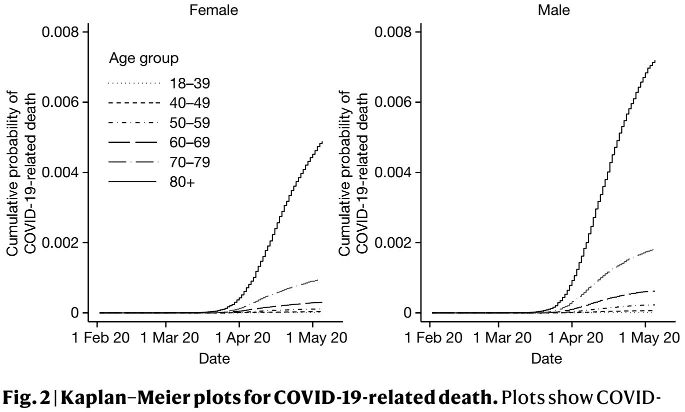


Fig. 2 | Kaplan–Meier plots for COVID-19-related death. Plots show COVID-19-related death over time by age and sex.

Table 2 | Hazard ratios and 95% confidence intervals for COVID-19-related death

Characteristic	Category	COVID-19 death HR (95% CI)	
		Adjusted for age and sex	Fully adjusted
Age	18–39	0.05 (0.04–0.07)	0.06 (0.04–0.08)
	40–49	0.28 (0.23–0.33)	0.30 (0.25–0.36)
	50–59	1.00 (ref)	1.00 (ref)
	60–69	2.79 (2.52–3.10)	2.40 (2.16–2.66)
	70–79	8.62 (7.84–9.46)	6.07 (5.51–6.69)
	80+	38.29 (35.02–41.87)	20.60 (18.70–22.68)
Sex	Female	1.00 (ref)	1.00 (ref)
	Male	1.78 (1.71–1.85)	1.59 (1.53–1.65)
BMI (kg m^{-2})	Not obese	1.00 (ref)	1.00 (ref)
	30–34.9 (obese class I)	1.23 (1.17–1.30)	1.05 (1.00–1.11)
	35–39.9 (obese class II)	1.81 (1.68–1.95)	1.40 (1.30–1.52)
	≥40 (obese class III)	2.66 (2.39–2.95)	1.92 (1.72–2.13)
Smoking	Never	1.00 (ref)	1.00 (ref)
	Former	1.43 (1.37–1.49)	1.19 (1.14–1.24)
	Current	1.14 (1.05–1.23)	0.89 (0.82–0.97)
Ethnicity[a]	White	1.00 (ref)	1.00 (ref)
	Mixed	1.62 (1.26–2.08)	1.43 (1.11–1.84)
	South Asian	1.69 (1.54–1.84)	1.45 (1.32–1.58)
	Black	1.88 (1.65–2.14)	1.48 (1.29–1.69)
	Other	1.37 (1.13–1.65)	1.33 (1.10–1.61)
IMD quintile	1 (least deprived)	1.00 (ref)	1.00 (ref)
	2	1.16 (1.08–1.23)	1.12 (1.05–1.19)
	3	1.31 (1.23–1.40)	1.22 (1.15–1.30)
	4	1.69 (1.59–1.79)	1.51 (1.42–1.61)
	5 (most deprived)	2.11 (1.98–2.25)	1.79 (1.68–1.91)
Blood pressure	Normal	1.00 (ref)	1.00 (ref)
	High blood pressure or diagnosed hypertension	1.09 (1.05–1.14)	0.89 (0.85–0.93)
Respiratory disease excluding asthma		1.95 (1.86–2.04)	1.63 (1.55–1.71)
Asthma[b] (versus none)	With no recent OCS use	1.13 (1.07–1.20)	0.99 (0.93–1.05)
	With recent OCS use	1.55 (1.39–1.73)	1.13 (1.01–1.26)
Chronic heart disease		1.57 (1.51–1.64)	1.17 (1.12–1.22)
Diabetes[c] (versus none)	With HbA1c < 58 mmol mol^{-1}	1.58 (1.51–1.66)	1.31 (1.24–1.37)
	With HbA1c ≥ 58 mmol mol^{-1}	2.61 (2.46–2.77)	1.95 (1.83–2.08)
	With no recent HbA1c measure	2.27 (2.06–2.50)	1.90 (1.72–2.09)
Cancer (non-haematological, versus none)	Diagnosed <1 year ago	1.81 (1.58–2.07)	1.72 (1.50–1.96)
	Diagnosed 1–4.9 years ago	1.20 (1.10–1.32)	1.15 (1.05–1.27)
	Diagnosed ≥5 years ago	0.99 (0.93–1.06)	0.96 (0.91–1.03)
Haematological malignancy (versus none)	Diagnosed <1 year ago	3.02 (2.24–4.08)	2.80 (2.08–3.78)
	Diagnosed 1–4.9 years ago	2.56 (2.14–3.06)	2.46 (2.06–2.95)
	Diagnosed ≥5 years ago	1.70 (1.46–1.98)	1.61 (1.39–1.87)
Reduced kidney function[d] (versus none)	eGFR 30–60	1.56 (1.49–1.63)	1.33 (1.28–1.40)
	eGFR < 30	3.48 (3.23–3.75)	2.52 (2.33–2.72)
Liver disease		2.39 (2.06–2.77)	1.75 (1.51–2.03)
Stroke or dementia		2.57 (2.46–2.70)	2.16 (2.06–2.27)
Other neurological disease		3.08 (2.85–3.33)	2.58 (2.38–2.79)
Organ transplant		6.00 (4.73–7.61)	3.53 (2.77–4.49)
Asplenia		1.62 (1.19–2.21)	1.34 (0.98–1.83)
Rheumatoid arthritis, lupus or psoriasis		1.30 (1.21–1.38)	1.19 (1.11–1.27)
Other immunosuppressive condition		2.75 (2.10–3.62)	2.21 (1.68–2.90)

Models were adjusted for age using a four-knot cubic spline for age, except for estimation of age-group hazard ratios. Ref, reference group; 95% CI, 95% confidence interval.

[a] Ethnicity hazard ratios were estimated from a model restricted to those with recorded ethnicity.

[b] For OCS use, 'recent' refers to during the year before baseline.

[c] Classification by HbA1c is based on measurements within 15 months of baseline.

[d] eGFR is measured in ml min^{-1} per 1.73 m^2 and taken from the most recent serum creatinine measurement.

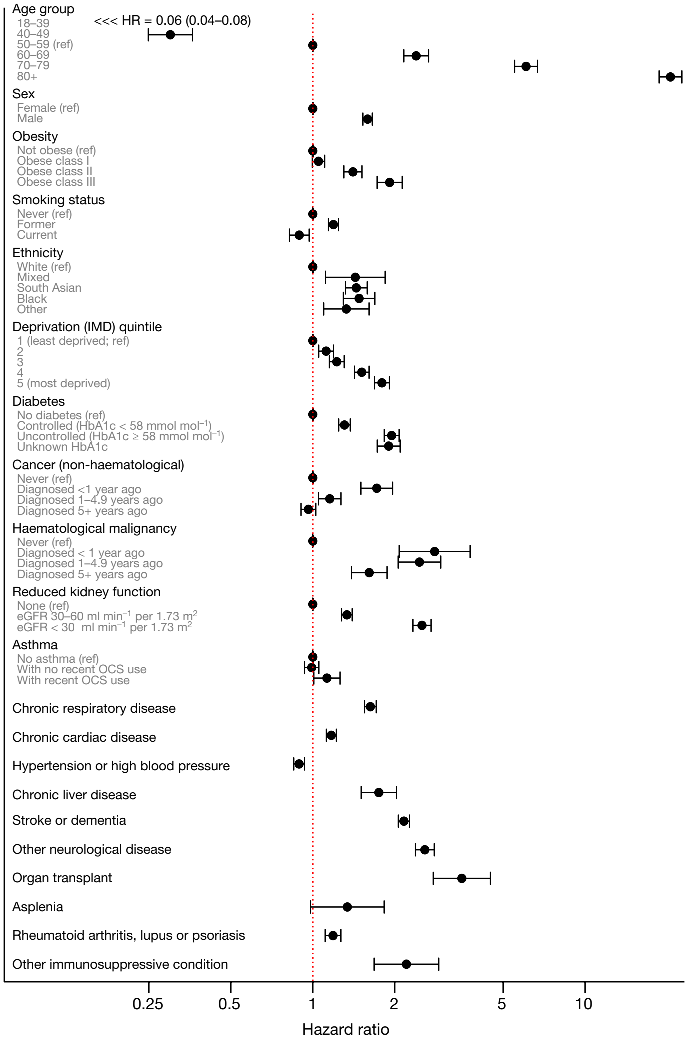


Fig. 3 | Estimated hazard ratios for each patient characteristic from a multivariable Cox model. Hazard ratios are shown on a log scale. Error bars represent the limits of the 95% confidence interval for the hazard ratio. IMD, index of multiple deprivation; obese class I, BMI 30–34.9; obese class II, BMI 35–39.9; obese class III, BMI ≥ 40; OCS, oral corticosteroid; ref, reference group. All hazard ratios are adjusted for all other factors listed other than ethnicity. Ethnicity estimates are from a separate model among those individuals for whom complete ethnicity data were available, and are fully adjusted for other covariates. Total $n = 17{,}278{,}392$ for the non-ethnicity models, and 12,718,279 for the ethnicity model.

resulted in some patients being incorrectly identified as having COVID-19. In addition, some COVID-19-related deaths may have been misclassified as non-COVID-19, particularly in the early stages of the pandemic; however, this inaccuracy is likely to have reduced quickly as the number of deaths increased, and a degree of outcome underascertainment—providing it is unrelated to patient characteristics—should not have biased our hazard ratios. Owing to the rarity of the outcome, the associations observed will be driven primarily by the profile of patient characteristics in the included cases. Our findings reflect both an individual's risk of infection and their risk of dying once infected. We will consider more detailed patient trajectories in future research within the OpenSAFELY platform.

Our large population may not be fully representative. We include only 17% of general practices in London—where many of the earlier cases of COVID-19 occurred—owing to the substantial geographical variation in the choice of electronic health record system. The user interface of electronic health records can affect prescribing of certain medicines[21–23], so it is possible that coding varies between systems.

Primary care records are detailed and longitudinal, but can be incomplete for data on patient characteristics. Ethnicity was missing for approximately 26% of patients, but was broadly representative[24]; there were also missing data on obesity and smoking. Sensitivity analyses found that our estimates were robust to our assumptions around missing data.

Non-proportional hazards could be due to very large numbers or unmeasured covariates. However, rapid changes in social behaviours (social distancing, shielding) and changes in the burden of infection may also have affected patient groups differentially. The larger hazard ratios seen for several covariates in a sensitivity analysis with earlier censoring (soon after social distancing and shielding policies were introduced) are consistent with patients who are more at risk being more compliant with these policies. By contrast, the risk associated with deprivation may have increased over time. Further analyses will explore the changes before and after the implementation of national initiatives around COVID-19.

Policy implications and interpretation

The UK has a policy of recommending shielding (staying at home at all times and avoiding any face-to-face contact) for groups who are identified as being extremely vulnerable to COVID-19 on the basis of pre-existing medical conditions[25]. We were able to evaluate the association between most of these conditions and death from COVID-19, and we confirmed the increased mortality risks, supporting the targeted use of additional protection measures for people in these groups. We have demonstrated that only a small part of the substantially increased risks of COVID-19-related death among BAME groups and among people living in more-deprived areas can be attributed to existing disease. Improved strategies to protect people in these groups are urgently needed[26]. These might include the specific consideration of BAME groups in shielding guidelines and workplace policies. Studies are needed to investigate the interplay of additional factors that we were unable to examine, including employment, access to personal protective equipment and the related risk of exposure to infection, and household density.

The UK has an unusually large volume of very detailed longitudinal patient data, especially through primary care, and we believe the UK has a responsibility to the global community to make good use of this data. OpenSAFELY demonstrates—on a very large scale—that this can be done securely, transparently and rapidly. We will enhance the OpenSAFELY platform to further inform the global response to the COVID-19 emergency.

Future research

The underlying causes of the higher risk of COVID-19-related death among BAME individuals, and among people from deprived areas, require further investigation. We would suggest collecting data on occupational exposure and living conditions as first steps. The statistical power offered by our approach means that associations with less-common factors can be robustly assessed in more detail and at the earliest possible date as the pandemic progresses. We will therefore update our findings and address smaller risk groups as new cases arise over time. The open source reusable codebase on OpenSAFELY supports the rapid, secure and collaborative development of new analyses; we are currently conducting expedited studies on the effects of various medical treatments and population interventions on the risk of COVID-19 infection, admission to intensive care units and death, alongside other observational analyses. OpenSAFELY is rapidly scalable for the incorporation of more NHS patient records, and new sources of data are progressing.

In conclusion, we have generated early insights into factors associated with COVID-19-related death using the detailed primary care records of 17 million NHS patients, while maintaining privacy, in the context of a global health emergency.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2521-4.

1. WHO. COVID-19 situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (2020).
2. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (2020).
3. Department of Social Care and Public Health England. Number of coronavirus (COVID-19) cases and risk in the UK. https://web.archive.org/web/20200501084711/https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public (2020).
4. NHS England. COVID-19 daily deaths. https://web.archive.org/web/20200501094237/https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/ (2020).
5. Deng, G., Yin, M., Chen, X. & Zeng, F. Clinical determinants for fatality of 44,672 patients with COVID-19. *Crit. Care* **24**, 179 (2020).
6. Docherty, A. B. et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO clinical characterisation protocol. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.23.20076042 (2020).
7. Simonnet, A. et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity* **28**, 1195–1199 (2020).
8. Lighter, J. et al. Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission. *Clin. Infect. Dis.* **2020**, ciaa415 (2020).
9. Simons, D., Shahab, L., Brown, J. & Perski, O. The association of smoking status with SARS-CoV-2 infection, hospitalisation and mortality from COVID-19: a living rapid evidence review. *Qeios* https://doi.org/10.32388/UJR2AW.2 (2020).
10. Guan, W.-J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
11. Miyara, M. et al. Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19. *Qeios* https://doi.org/10.32388/WPP19W.3 (2020).
12. Khunti, K., Singh, A. K., Pareek, M. & Hanif, W. Is ethnicity linked to incidence or outcomes of covid-19? *Br. Med. J.* **369**, m1548 (2020).
13. Institute for Fiscal Studies. Are some ethnic groups more vulnerable to COVID-19 than others? https://web.archive.org/web/20200502130148/https://www.ifs.org.uk/inequality/chapter/are-some-ethnic-groups-more-vulnerable-to-covid-19-than-others/ (2020).
14. Public Health England. Disparities in the risk and outcomes of COVID-19. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/889195/disparities_review.pdf (2020).
15. Intensive Care National Audit and Research Centre. COVID-19 report. https://web.archive.org/web/20200425133758/https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports (2020).
16. Office for National Statistics. Deaths registered weekly in England and Wales, provisional: week ending 17 April 2020. https://web.archive.org/web/20200430191844/https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending17april2020 (2020).
17. Halpin, D. M. G., Faner, R., Sibila, O., Badia, J. R. & Agusti, A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection? *Lancet Respir. Med.* **8**, 436–438 (2020).
18. Boddington, N. L. et al. COVID-19 in Great Britain: epidemiological and clinical characteristics of the first few hundred (FF100) cases: a descriptive case series and case control analysis. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.18.20086157 (2020).
19. Rentsch, C. T. et al. Covid-19 testing, hospital admission, and intensive care among 2,026,227 United States veterans aged 54–75 years. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.09.20059964 (2020).
20. Farsalinos, K., Barbouni, A. & Niaura, R. Smoking, vaping and hospitalization for COVID-19. *Qeios* https://doi.org/10.32388/Z69O8A.13 (2020).
21. MacKenna, B. et al. Impact of electronic health record interface design on unsafe prescribing of ciclosporin, tacrolimus and diltiazem: a cohort study in English NHS primary care. *J. Med. Internet Res.* https://doi.org/10.2196/17003 (2020).
22. Opondo, D. et al. Quality of co-prescribing NSAID and gastroprotective medications for elders in The Netherlands and its association with the electronic medical record. *PLoS One* **10**, e0129515 (2015).
23. MacKenna, B. Ghost branded generics: why does the cost of generic atorvastatin vary? *EBM DataLab* https://web.archive.org/web/20200502135915/https://ebmdatalab.net/ghost-branded-generics-why-does-the-cost-of-generic-atorvastatin-vary%ef%bb%bf/ (2018).
24. Mathur, R. et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J. Public Health* **36**, 684–692 (2014).
25. Public Health England. Guidance on shielding and protecting people who are clinically extremely vulnerable from COVID-19. https://web.archive.org/web/20200501090127/https://www.gov.uk/government/publications/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19 (2020).
26. Marmot, M., Friel, S., Bell, R., Houweling, T. A. & Taylor, S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* **372**, 1661–1669 (2008).

Methods

Study design

We conducted a cohort study using national primary care electronic health record data linked to data on COVID-19-related deaths (see 'Data source'). The cohort study began on 1 February 2020, which was chosen as a date several weeks before the first reported COVID-19-related deaths and the day after the second laboratory-confirmed case[27]; and ended on 6 May 2020. The cohort study examines risk among the general population rather than in a population infected with SARS-COV-2. Therefore, all patients were included irrespective of any SARS-COV-2 test results. No randomization was undertaken. Outcome assessment was undertaken as part of routine health care, therefore no blinding of any sort was attempted. However, study investigators had no involvement in outcome assessment.

Data source

We used patient data from general practice (GP) records managed by the GP software provider The Phoenix Partnership (TPP), linked to death data from the ONS. ONS data include information on all deaths, including COVID-19-related death (defined as a COVID-19 ICD-10 code mentioned anywhere on the death certificate) and non-COVID-19 death, which was used for censoring.

The data were accessed, linked and analysed using OpenSAFELY, a new data analytics platform that was created to address urgent questions relating to the epidemiology and treatment of COVID-19 in England. OpenSAFELY provides a secure software interface that allows detailed pseudonymized primary care patient records to be analysed in near-real time where they already reside—hosted within the highly secure data centre of the electronic health records vendor—to minimize the reidentification risks when data are transported off-site; other smaller datasets are linked to these data within the same environment using a matching pseudonym derived from the NHS number. More information can be found at https://opensafely.org/.

The dataset that was analysed with OpenSAFELY is based on around 24 million currently registered patients (approximately 40% of the English population) from GP surgeries using the TPP SystmOne electronic health record system. SystmOne is a secure centralized electronic health records system that has been used in English clinical practice since 1998; it records data entered (in real time) by GPs and practice staff during routine primary care. The system is accredited under the NHS-approved systems framework for general practice[28,29]. Data extracted from TPP SystmOne have previously been used in medical research, as part of the ResearchOne dataset[30,31]. From these electronic health records a pseudonymized dataset was created for OpenSAFELY that consisted of 20 billion rows of structured data; including, for example, the diagnoses, medications, physiological parameters and prior investigations of pseudonymized patients (Extended Data Fig. 2, level 1). All OpenSAFELY data processing took place on TPP's servers; external data providers securely transferred pseudoymized data (such as COVID-19-related death from ONS) for linkage to OpenSAFELY (Extended Data Fig. 2, level 2); and study definitions developed in Python on GitHub were pulled into the OpenSAFELY infrastructure and used to create a study dataset of one row per patient (Extended Data Fig. 2, level 3). Statistical code was developed using synthetic data and used to analyse the study dataset; this included code to check data ranges, to check consistency of data columns and to produce descriptive statistics for comparison with expected disease prevalences to ensure validity, as well as code to fit our analysis models. Only two authors (K.B. and A.J.W.) accessed OpenSAFELY to run code; no pseudonymized patient-level data were ever removed from TPP infrastructure; and only aggregated, anonymous, manually checked study results were released for publication (Extended Data Fig. 2, level 4), All code for data management and analysis is archived online (see 'Code availability').

Study population and observation period

Our study population consisted of all adults (males and females 18 years and above) currently registered as active patients in a TPP GP surgery in England on 1 February 2020. To be included in the study, participants were required to have at least one year of prior follow-up in the GP practice to ensure that baseline patient characteristics could be adequately captured, and to have recorded sex, age and deprivation[32] (see 'Covariates'). Patients were observed from 1 February 2020 and were followed until the first of either their death date (whether COVID-19-related or due to other causes) or the study end date, 6 May 2020. For this analysis, ONS death data were available to 11 May 2020, but we used an earlier censor date to allow for delays in reporting of the last few days of available data.

Outcomes

The outcome was COVID-19-related death; this was ascertained from ONS death certificate data in which the COVID related ICD-10 codes U071 or U072 were present in the record.

Covariates

Characteristics included: health conditions listed in UK guidance on 'higher risk' groups[33]; other common conditions that may cause immunodeficiency inherently or through medication (cancer and common autoimmune conditions); and emerging risk factors for severe outcomes among COVID-19 cases (such as raised blood pressure).

Age, sex, BMI (kg m^{-2}) and smoking status were included. Where categorized, age groups were: 18–39, 40–49, 50–59, 60–69, 70–79 and 80+ years. BMI was ascertained from weight measurements within the last 10 years, restricted to those taken when the patient was over 16 years old. Obesity was grouped using categories derived from the WHO classification of BMI: no evidence of obesity, BMI < 30; obese class I, BMI 30–34.9; obese class II, BMI 35–39.9; and obese class III, BMI 40+. Smoking status was grouped into current-, former- and never-smokers.

The following comorbidities were also considered: asthma, other chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, chronic neurological diseases, common autoimmune diseases (rheumatoid arthritis, systemic lupus erythematosus or psoriasis), solid organ transplant, asplenia, other immunosuppressive conditions, cancer, evidence of reduced kidney function, and raised blood pressure or a diagnosis of hypertension.

Disease groupings followed national guidance on risk of influenza infection[34], therefore 'chronic respiratory disease (other than asthma)' included chronic obstructive pulmonary disease, fibrosing lung disease, bronchiectasis or cystic fibrosis; and 'chronic heart disease' included chronic heart failure, ischaemic heart disease, and severe valve or congenital heart disease likely to require lifelong follow-up. Chronic neurological conditions were separated into diseases with a probable cardiovascular aetiology (stroke, transient ischaemic attack, dementia) and conditions in which respiratory function may be compromised, such as motor neurone disease, myasthenia gravis, multiple sclerosis, Parkinson's disease, cerebral palsy, quadriplegia or hemiplegia and progressive cerebellar disease. Asplenia included splenectomy or a spleen dysfunction, including sickle cell disease. Other immunosuppressive conditions included human immunodeficiency virus (HIV) or a condition inducing permanent immunodeficiency ever diagnosed, or aplastic anaemia or temporary immunodeficiency recorded within the last year. Haematological malignancies were considered separately from other cancers to reflect the immunosuppression associated with haematological malignancies and their treatment. Kidney function was ascertained from the most recent serum creatinine measurement, where available, and was converted into the eGFR using the chronic kidney disease epidemiology collaboration (CKD-EPI) equation[35], with reduced kidney function grouped into eGFR 30–59.9 or <30 ml min^{-1} per 1.73 m^2. History of kidney dialysis or end-stage renal failure was separately explored in a secondary analysis. Raised blood pressure was defined as either a previous coded diagnosis of hypertension or the most recent recording indicating systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg.

Asthma was grouped by use of oral corticosteroids as an indication of severity. Diabetes was grouped according to the most recent Hba1c measurement within the last 15 months (Hba1c < 58 mmol mol^{-1}; Hba1c ≥ 58 mmol mol^{-1}; or no recent measure available). Cancer was grouped by time since the first diagnosis (within the last year; between 1 and 4.9 years ago; more than 5 years ago).

Other covariates that were considered as potential upstream factors were deprivation and ethnicity. Deprivation was measured by the index of multiple deprivation (IMD, in quintiles, with higher values indicating greater deprivation), derived from the patient's postcode at lower super output area level for a high degree of precision. Ethnicity was grouped into white, Black, South Asian, mixed, or other. In sensitivity analyses, a more detailed grouping of ethnicity was explored. The Sustainability and Transformation Partnership (STP, an NHS administrative region) of the patient's general practice was included as an additional adjustment for geographical variation in infection rates across the country.

Information on all covariates was obtained from primary care records by searching TPP SystmOne records for specific coded data. TPP SystmOne allows users to work with the SNOMED-CT clinical terminology, using a GP subset of SNOMED-CT codes. This subset maps on to the native Read version 3 (CTV3) clinical coding system on which SystmOne is built. Medicines are entered or prescribed in a format compliant with the NHS Dictionary of Medicines and Devices (dm+d)[36], a local UK extension library of SNOMED. Codelists for particular underlying conditions and medicines were compiled from a variety of sources. These include British National Formulary (BNF) codes from OpenPrescribing.net, published codelists for asthma[37–39], immunosuppression[40–42], psoriasis[43], systemic lupus erythematosus[44], rheumatoid arthritis[45,46] and cancer[47,48], and Read Code 2 lists designed specifically to describe groups who are at increased risk of influenza infection[18]. Read Code 2 lists were added to with SNOMED codes and cross-checked against NHS Quality and Outcomes Framework (QOF) registers, then translated into CTV3 with manual curation. Decisions on every codelist were documented and the final lists were reviewed by at least two authors. Detailed information on compilation and sources for every individual codelist is available at https://codelists.opensafely.org/ and the lists are available for inspection and reuse by the broader research community.

Statistical analysis

Patient numbers are depicted in a flowchart (Fig. 1). The Kaplan–Meier failure function was estimated by age group and sex. For each patient characteristic, a Cox proportional hazards model was fitted, with days in study as the timescale, stratified by geographical area (STP), and adjusted for sex and age modelled using restricted cubic splines. Violations of the proportional hazards assumption were explored by testing for a zero slope in the scaled Schoenfeld residuals. All patient characteristics, including age (again modelled as a spline), sex, BMI, smoking, IMD quintile, and comorbidities listed above were then included in a single multivariable Cox proportional hazards model, stratified by STP. Hazard ratios from the age-and-sex adjusted and fully adjusted models are reported with 95% confidence intervals. Models were also refitted with age group fitted as a categorical variable to obtain hazard ratios by age group.

In the primary analysis, those with missing BMI were assumed to be non-obese and those with missing smoking information were assumed to be non-smokers on the assumption that both obesity and smoking would be likely to be recorded if present. A sensitivity analysis was run among those with complete BMI and smoking data only. Ethnicity was omitted from the main multivariable model owing data being missing for 26% of individuals; hazard ratios for ethnicity were therefore obtained from a separate model among individuals with complete ethnicity data only. Hazard ratios for other patient characteristics, adjusted for ethnicity, were also obtained from this model and are presented in the sensitivity analyses to allow assessment of whether estimates were distorted by ethnicity in the primary model. We conducted an additional sensitivity analysis using a population-calibrated imputation approach to handle missing ethnicity[49,50], with marginal proportions of each ethnicity group within each of nine broad geographical regions of England (East, East Midlands, London, North East, North West, South East, South West, West Midlands, Yorkshire and The Humber) taken from Annual Population Survey (APS) data (pooled 2014–2016)[51]. Five imputed datasets were created with estimated hazard ratios combined using Rubin's rules.

The C-statistic was calculated as a measure of model discrimination. Owing to computational time, this was estimated by randomly sampling 5,000 patients with and without the outcome and calculating the C-statistic using the random sample, repeating this 10 times and taking the average C-statistic. Weights were applied to account for the sampling[56].

All *P* values presented are two-sided.

Information governance and ethics

NHS England is the data controller; TPP is the data processor; and the key researchers on OpenSAFELY are acting on behalf of NHS England. This implementation of OpenSAFELY is hosted within the TPP environment, which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant[52,53]; patient data have been pseudonymized for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymized datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the platform is through a virtual private network (VPN) connection, restricted to a small group of researchers, their specific machine and IP address; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; and only aggregate statistical outputs leave the platform environment following best practice for anonymization of results such as statistical disclosure control for low cell counts[54]. The OpenSAFELY research platform adheres to the data protection principles of the UK Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR) 2016. In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organizations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure[55]. Together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. GP practices, from which the primary care data are obtained, are required to share relevant health information to support the public health response to the pandemic, and have been informed of the OpenSAFELY analytics platform. This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the London School of Hygiene and Tropical Medicine (LSHTM) ethics board (reference 21863). No further ethical or research governance approval was required by the University of Oxford but copies of the approval documents were reviewed and held on record. Guarantor: B.G. and L.S.

Patient and public involvement

Patients were not formally involved in developing this specific study design. We have developed a publicly available website (https://opensafely.org/) that allows any patient or member of the public to contact us regarding this study or the broader OpenSAFELY project. This feedback will be used to refine and prioritize our OpenSAFELY activities.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data were linked, stored and analysed securely within the OpenSAFELY platform (https://opensafely.org/). Detailed pseudonymized patient data are potentially reidentifiable and therefore not shared. We rapidly delivered the OpenSAFELY data analysis platform without prior funding to deliver timely analyses on urgent research questions in the context of the global COVID-19 health emergency: now that the platform is established we are developing a formal process for external users to request access in collaboration with NHS England. Details of this process will be published shortly on the OpenSAFELY website.

Code availability

Data management was performed using Python 3.8 and SQL, with analysis carried out using Stata 16.1 and Python. All code is shared openly for review and reuse under an MIT open license. All code for data management and analysis is archived online at https://github.com/opensafely/risk-factors-research. All clinical and medicines codelists are openly available for inspection and reuse at https://codelists.opensafely.org/.

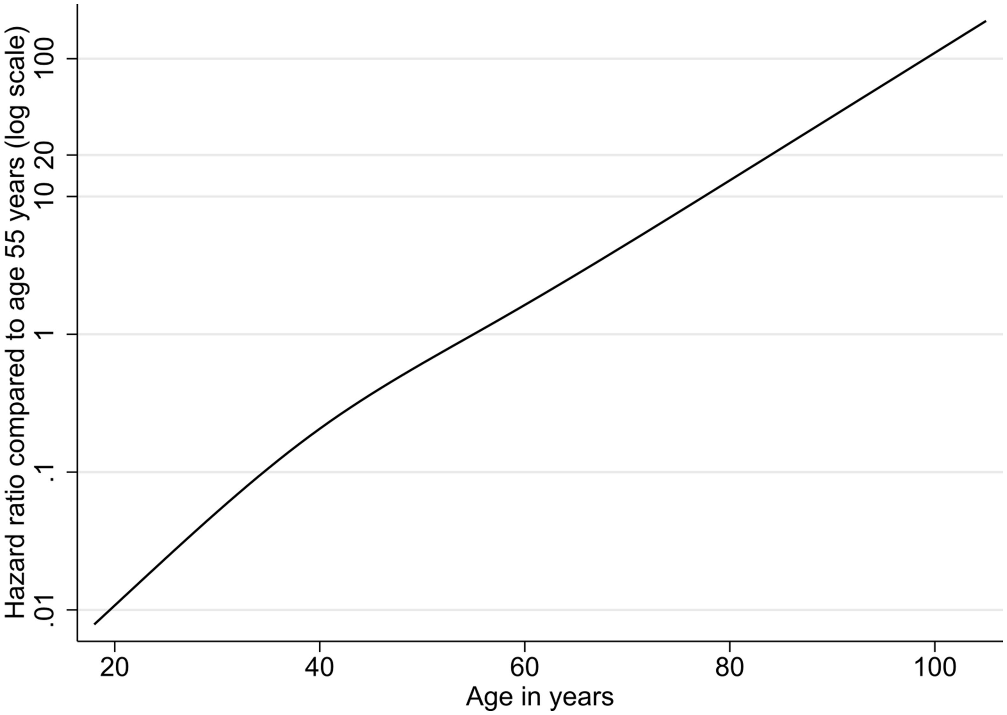
27. UK Government. Coronavirus (COVID-19) cases in the UK. https://web.archive.org/web/20200502045059/https://coronavirus.data.gov.uk/ (2020).
28. NHS Digital. GP systems of choice. https://digital.nhs.uk/services/gp-systems-of-choice (2020).
29. NHS Digital. Future GP IT systems and services. https://digital.nhs.uk/services/future-gp-it-systems-and-services (2020).
30. Clegg, A. et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* **45**, 353–360 (2016).
31. Harcourt, S. et al. Estimating primary care attendance rates for fever in infants after meningococcal B vaccination in England using national syndromic surveillance data. *Vaccine* **36**, 565–571 (2018).
32. Lewis, J. D., Bilker, W. B., Weinstein, R. B. & Strom, B. L. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol. Drug Saf*. **14**, 443–451 (2005).
33. Public Health England. Guidance on social distancing for everyone in the UK. https://web.archive.org/web/20200429043059/https://www.gov.uk/government/publications/covid-19-guidance-on-social-distancing-and-for-vulnerable-people/guidance-on-social-distancing-for-everyone-in-the-uk-and-protecting-older-people-and-vulnerable-adults (2020).
34. Public Health England. UK immunisation schedule: the green book, chapter 11. https://www.gov.uk/government/publications/immunisation-schedule-the-green-book-chapter-11 (2013).
35. Levey, A. S. et al. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med*. **150**, 604–612 (2009).
36. MacKenna, B. What is the dm+d? The NHS Dictionary of Medicines and Devices. *EBM DataLab* https://web.archive.org/web/20200502143707/https://ebmdatalab.net/what-is-the-dmd-the-nhs-dictionary-of-medicines-and-devices/ (2019).
37. Nissen, F. et al. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). *BMJ Open* **7**, e017474 (2017).
38. Morton, C. & Douglas, I. OpenSAFELY codelists: asthma diagnosis. https://codelists.opensafely.org/codelist/opensafely/asthma-diagnosis/ (2020).
39. MacKenna, B. & Douglas, I. OpenSAFELY codelists: asthma oral prednisolone medication. https://codelists.opensafely.org/codelist/opensafely/asthma-oral-prednisolone-medication/ (2020).
40. Grint, D. J. et al. Safety of inadvertent administration of live zoster vaccine to immunosuppressed individuals in a UK-based observational cohort analysis. *BMJ Open* **10**, e034886 (2020).
41. McDonald, H. & Smeeth, L. OpenSAFELY codelists: permanent immunosuppression. https://codelists.opensafely.org/codelist/opensafely/permanent-immunosuppression/ (2020).
42. Smeeth, L. & McDonald, H. OpenSAFELY codelists: temporary immunosuppression. https://codelists.opensafely.org/codelist/opensafely/temporary-immunosuppression/ (2020).
43. Wong, A., Schmidt, S. A. J. & Langan, S. Clinical code list – psoriasis – read codes [Data collection]. https://doi.org/10.17037/DATA.00001255 (London School of Hygiene and Tropical Medicine, 2019).
44. Forbes, H. et al. Clinical code list – SLE codes [Data collection]. https://doi.org/10.17037/DATA.162 (London School of Hygiene and Tropical Medicine, 2014).
45. Pujades-Rodriguez, M. et al. Rheumatoid arthritis and incidence of twelve initial presentations of cardiovascular disease: a population record-linkage cohort study in England. *PLoS One* **11**, e0151245 (2016).
46. Morton, C. & Tomlinson, L. Open SAFELY codelists: RA/SLE/psoriasis. https://codelists.opensafely.org/codelist/opensafely/ra-sle-psoriasis/ (2020).
47. Strongman, H. et al. Medium and long-term risks of specific cardiovascular diseases in survivors of 20 adult cancers: a population-based cohort study using multiple linked UK electronic health records databases. *Lancet* **394**, 1041–1054 (2019).
48. Morton, C. & Walker, A. Open SAFELY codelists: cancer excluding lung and haematological. https://codelists.opensafely.org/codelist/opensafely/cancer-excluding-lung-and-haematological/ (2020).
49. Carpenter, J. R. & Kenward, M. G. *Multiple Imputation and its Application* (John Wiley & Sons, 2012).
50. Pham, T. M., Carpenter, J. R., Morris, T. P., Wood, A. M. & Petersen, I. Population-calibrated multiple imputation for a binary/categorical covariate in categorical regression models. *Stat. Med*. **38**, 792–808 (2019).
51. Office for National Statistics. Population characteristics research tables. https://web.archive.org/web/20200513113451/https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationcharacteristicsresearchtables (2019).
52. NHS Digital. BETA – data security standards. https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework/beta---data-security-standards (2020).
53. NHS Digital. Data security and protection toolkit. https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit (2018).
54. NHS Digital. ISB1523: Anonymisation standard for publishing health and social care data. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1523-anonymisation-standard-for-publishing-health-and-social-care-data (2019).
55. Department of Health and Social Care. Coronavirus (COVID-19): notification to organisations to share information. https://web.archive.org/web/20200421171727/https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information (2020).
56. Sanderson, J., Thompson, S.G., White, I.R., Aspelund, T. & Pennells, L. Derivation and assessment of risk prediction models using case-cohort data. *BMC Med. Res. Methodol*. **13**, 113 (2013).

Acknowledgements All authors are from The OpenSAFELY Collaborative. We are grateful for all the support received from the TPP Technical Operations team throughout this work; for assistance from the information governance and database teams at NHS England and NHSX; and for additional discussions on disease characterization, codelists and methodology with H. Drysdale, B. Nicholson, N. DeVito, W. Hulme, I. Lipska, J. Morley, J. Quint and T. Pham. No dedicated funding has yet been obtained for this work. TPP provided technical expertise and infrastructure within their data centre pro bono in the context of a national emergency. The work of B.G. on better use of data in healthcare more broadly is currently funded in part by: the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, NIHR Applied Research Collaboration Oxford and Thames Valley, the Mohn-Westlake Foundation, NHS England and the Health Foundation; all DataLab staff are supported by the grants of B.G. for this work. L.S. reports grants from Wellcome, MRC, NIHR, UKRI, British Council, GSK, British Heart Foundation and Diabetes UK outside this work; K.B. holds a Sir Henry Dale fellowship jointly funded by Wellcome and the Royal Society; H.I.M. is funded by the NIHR Health Protection Research Unit in Immunisation (a partnership between Public Health England and LSHTM); A.Y.S.W. holds a fellowship from BHF; R.M. holds a Sir Henry Wellcome fellowship funded by the Wellcome Trust; E.J.W. holds grants from MRC; R.G. holds grants from NIHR and MRC; I.J.D. holds grants from NIHR and GSK; and H.F. holds a UKRI fellowship. The views expressed are those of the authors and not necessarily those of the NIHR, NHS England, Public Health England or the Department of Health and Social Care. The funders had no role in the study design; the collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

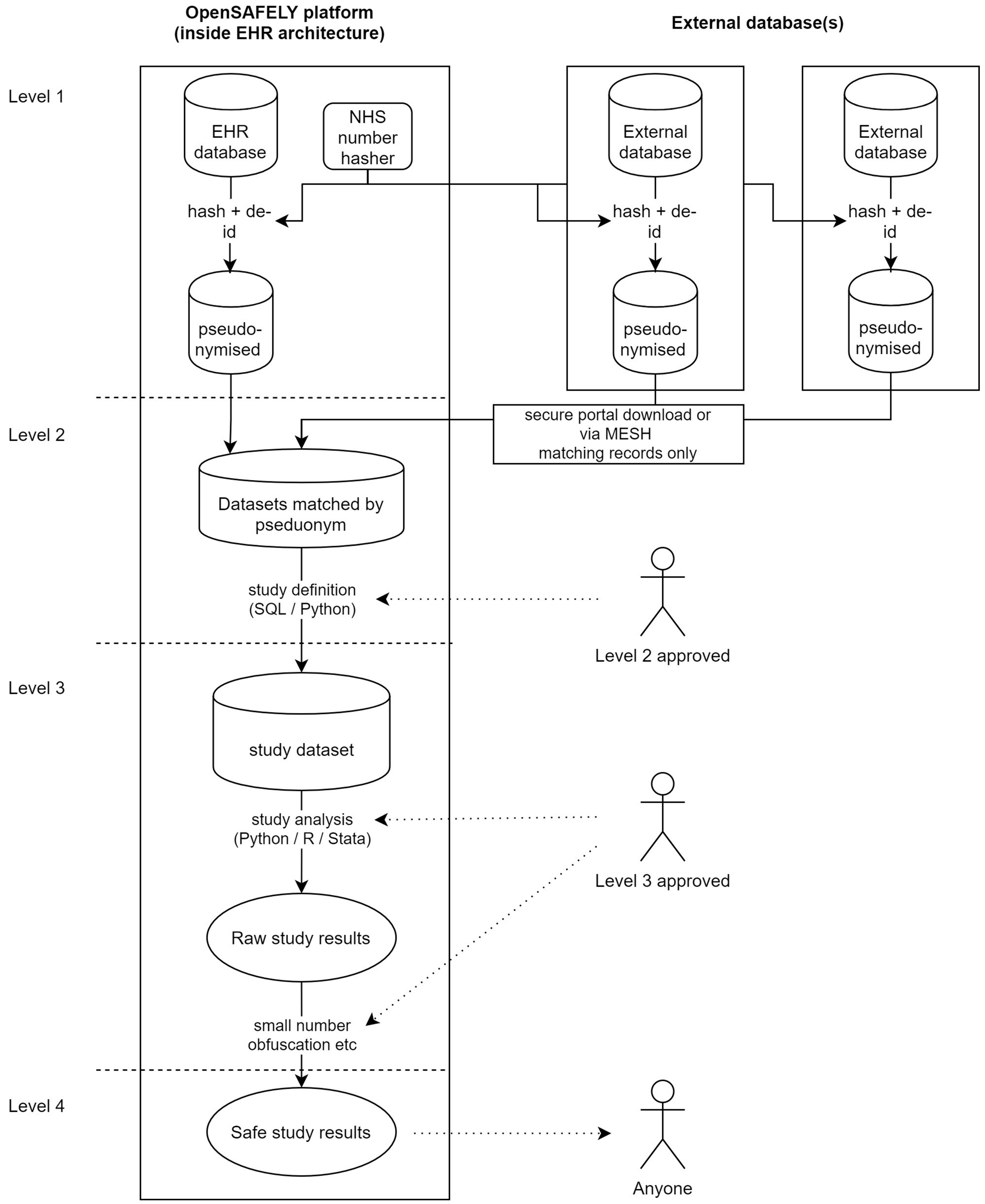
Author contributions B.G. conceived the platform and the approach; B.G. and L.S. led the project overall and are guarantors; S.B. led the software; E.J.W and K.B. led the statistical analysis; C.E.M. and A.J.W. led on codelists and implementation; and A.M. led on information governance. Contributions are as follows: data curation, C.B., J.P., J.C., S.H., S.B., D.E., P.I. and C.E.M.; analysis, E.J.W., K.B., A.J.W. and C.E.M.; funding acquisition, B.G. and L.S.; information governance, A.M., B.G., C.B. and J.P.; methodology, E.J.W., K.B., A.J.W., B.G., L.S., C.B., J.P., J.C., S.H., S.B., D.E., P.I., C.E.M., R.G., D.H. and R.P.; disease category conceptualization and codelists, C.E.M., A.J.W., P.I., S.B., D.E., C.B., J.C., J.P., S.H., H.J.C., K.B., S.B., A.M., B.M., L.T., I.J.D., H.I.M., R.M. and H.F.; ethics approval, H.J.C., E.J.W., L.S. and B.G.; project administration, C.E.M., H.J.C., C.B., S.B., A.M., L.S. and B.G.; resources, B.G., L.S. and F.H.; software, S.B., D.E., P.I., A.J.W., C.E.M., C.B., F.H., J.C. and S.H.; supervision, B.G., L.S. and S.B.; writing (original draft), H.J.C., E.J.W., K.B., B.M., C.E.M., A.M., B.G. and L.S.; and writing (review and editing), C.B., C.E.M., H.J.C., E.J.W., K.B., S.B., A.M., B.M., L.T., I.J.D., H.I.M., R.M., A.J.W. and S.J.W.E. All authors were involved in design and conceptual development and reviewed and approved the final manuscript.

Competing interests All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form at www.icmje.org/coi_disclosure.pdf. C.B., J.P., F.H., J.C. and S.H. are employees of TPP. A.M. was interim Chief Medical Officer of NHS Digital April–Sept 2019 (left NHS Digital at the end of January 2020) and Digital Clinical Champion NHS England 2014–2015. All other authors have no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2521-4.
Correspondence and requests for materials should be addressed to B.G.
Peer review information *Nature* thanks David Christiani, Jeffrey Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Estimated log-transformed hazard ratio by age in years. From the primary fully adjusted model containing a four-knot cubic spline for age, and adjusted for all covariates listed in Table 2 except for ethnicity.



Extended Data Fig. 2 | Illustration of data flows in the OpenSAFELY platform. Overview of the architecture of the OpenSAFELY platform. EHR, electronic health record.

Extended Data Table 1 | Adjusted hazard ratios for detailed ethnicity categories

Ethnicity	Fully adjusted hazard ratio*	95% CI
British or mixed British	1.00	(ref)
Irish	1.16	(0.96-1.41)
Other White	0.87	(0.79-0.97)
Mixed ethnicity	1.42	(1.11-1.83)
Indian or British Indian	1.40	(1.23-1.59)
Pakistani or British Pakistani	1.24	(1.06-1.46)
Bangladeshi or British Bangladeshi	1.84	(1.36-2.49)
Other Asian	1.73	(1.44-2.09)
Caribbean	1.28	(1.07-1.53)
African	1.77	(1.41-2.22)
Other Black	1.73	(1.24-2.40)
Chinese	1.22	(0.81-1.85)
Other	1.35	(1.09-1.67)

Estimated from a model restricted to those with recorded ethnicity, adjusted for age (using a four-knot cubic spline for age), sex, BMI, smoking, IMD quintile, hypertension or high blood pressure, asthma, chronic heart disease, diabetes, non-haematological cancer, haematological malignancy, reduced kidney function, liver disease, stroke or dementia, other neurological disease, organ transplant, asplenia, rheumatoid arthritis, lupus or psoriasis, and other immunosuppressive condition. All categorizations are as in the primary analysis.

Extended Data Table 2 | Hazard ratios and 95% confidence intervals in sensitivity analyses

		Fully adjusted HR and 95% CI				
Characteristic	**Category**	**Primary analysis**	**Early censoring at 6/4/2020**	**Restricted to those with complete BMI /smoking**	**Adjusted for ethnicity in those where recorded**	**Adjusted for ethnicity using multiple imputation**
N outcome events in analysis		10926	2816	9880	8149	
Age	18-<40	0.06 (0.04-0.08)	0.07 (0.04-0.12)	0.07 (0.05-0.10)	0.07 (0.05-0.09)	0.06 (0.04-0.07)
	40-<50	0.30 (0.25-0.36)	0.33 (0.23-0.46)	0.30 (0.25-0.37)	0.29 (0.24-0.36)	0.29 (0.24-0.35)
	50-<60	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	60-<70	2.40 (2.16-2.66)	2.55 (2.11-3.08)	2.38 (2.13-2.66)	2.37 (2.11-2.67)	2.43 (2.19-2.70)
	70-<80	6.07 (5.51-6.69)	5.84 (4.88-6.98)	5.96 (5.37-6.61)	6.05 (5.42-6.76)	6.24 (5.66-6.87)
	80+	20.60 (18.70-22.68)	14.66 (12.23-17.58)	19.96 (18.00-22.14)	20.19 (18.08-22.54)	21.19 (19.23-23.34)
Sex	Female	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Male	1.59 (1.53-1.65)	1.89 (1.75-2.05)	1.65 (1.58-1.72)	1.54 (1.47-1.61)	1.57 (1.52-1.64)
BMI	Not obese	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	
	30-34.9kg/m2 (Obese class I)	1.05 (1.00-1.11)	1.30 (1.18-1.43)	1.07 (1.02-1.13)	1.05 (0.99-1.11)	1.06 (1.00-1.11)
	35-39.9kg/m2 (Obese class II)	1.40 (1.30-1.52)	1.57 (1.36-1.81)	1.45 (1.34-1.57)	1.41 (1.30-1.54)	1.42 (1.32-1.54)
	≥40 kg/m2 (Obese class III)	1.92 (1.72-2.13)	2.70 (2.26-3.21)	1.99 (1.79-2.21)	1.92 (1.70-2.17)	1.96 (1.76-2.18)
Smoking	Never	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Former	1.19 (1.14-1.24)	1.27 (1.17-1.39)	1.18 (1.13-1.24)	1.22 (1.16-1.29)	1.23 (1.18-1.29)
	Current	0.89 (0.82-0.97)	0.93 (0.79-1.09)	0.91 (0.83-0.99)	0.93 (0.84-1.02)	0.93 (0.85-1.01)
Ethnicity[a]	White	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Mixed	1.43 (1.11-1.84)	1.01 (0.60-1.72)	1.38 (1.05-1.80)	1.43 (1.11-1.84)	1.44 (1.06-1.95)
	South Asian	1.45 (1.32-1.58)	1.63 (1.38-1.91)	1.51 (1.38-1.66)	1.45 (1.32-1.58)	1.48 (1.33-1.65)
	Black	1.48 (1.29-1.69)	1.76 (1.41-2.19)	1.47 (1.28-1.69)	1.48 (1.29-1.69)	1.53 (1.32-1.77)
	Other	1.33 (1.10-1.61)	1.84 (1.37-2.47)	1.40 (1.15-1.71)	1.33 (1.10-1.61)	1.34 (1.12 1.61)
IMD quintile	1 (least deprived)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	2	1.12 (1.05-1.19)	0.96 (0.85-1.08)	1.12 (1.05-1.19)	1.16 (1.08-1.25)	1.12 (1.05-1.19)
	3	1.22 (1.15-1.30)	1.00 (0.88-1.12)	1.23 (1.15-1.31)	1.26 (1.17-1.36)	1.21 (1.14-1.29)
	4	1.51 (1.42-1.61)	1.26 (1.11-1.41)	1.51 (1.42-1.61)	1.54 (1.43-1.66)	1.48 (1.39-1.57)
	5 (most deprived)	1.79 (1.68-1.91)	1.41 (1.25-1.60)	1.80 (1.68-1.93)	1.77 (1.64-1.91)	1.72 (1.61-1.84)
Blood pressure	Normal	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	High bp or diagnosed hyper-tension	0.89 (0.85-0.93)	0.95 (0.87-1.04)	0.88 (0.84-0.92)	0.91 (0.86-0.96)	0.89 (0.85-0.93)
Respiratory disease ex asthma		1.63 (1.55-1.71)	1.86 (1.69-2.04)	1.59 (1.51-1.67)	1.65 (1.56-1.75)	1.64 (1.56-1.72)
Asthma (vs none)[b]	With no recent OCS use	0.99 (0.93-1.05)	1.08 (0.96-1.20)	0.97 (0.91-1.04)	0.94 (0.87-1.00)	0.98 (0.93-1.05)
	With recent OCS use	1.13 (1.01-1.26)	1.38 (1.13-1.67)	1.09 (0.97-1.22)	1.08 (0.95-1.23)	1.11 (0.99-1.24)
Chronic heart disease		1.17 (1.12-1.22)	1.37 (1.26-1.48)	1.16 (1.11-1.22)	1.16 (1.11-1.22)	1.17 (1.12-1.22)
Diabetes (vs none)[c]	With HbA1c<58 mmol/mol	1.31 (1.24-1.37)	1.39 (1.26-1.52)	1.29 (1.23-1.36)	1.28 (1.21-1.36)	1.27 (1.21-1.33)
	WithHbA1c>=58 mmol/mol	1.95 (1.83-2.08)	2.33 (2.08-2.61)	1.90 (1.78-2.02)	1.86 (1.73-2.00)	1.87 (1.76-1.99)
	With no recent HbA1c measure	1.90 (1.72-2.09)	1.71 (1.40-2.08)	1.92 (1.74-2.12)	1.86 (1.67-2.08)	1.84 (1.67-2.02)
Cancer (non-haematological, vs none)	Diagnosed < 1 year ago	1.72 (1.50-1.96)	1.66 (1.27-2.16)	1.68 (1.46-1.94)	1.67 (1.43-1.96)	1.74 (1.52-1.99)
	Diagnosed 1-4.9 years ago	1.15 (1.05-1.27)	1.34 (1.13-1.60)	1.16 (1.05-1.28)	1.21 (1.09-1.35)	1.17 (1.06-1.28)
	Diagnosed ≥5 years ago	0.96 (0.91-1.03)	0.92 (0.81-1.04)	0.97 (0.91-1.03)	0.98 (0.92-1.06)	0.97 (0.92-1.04)
Haematological malignancy (vs none)	Diagnosed < 1 year ago	2.80 (2.08-3.78)	2.20 (1.14-4.24)	2.86 (2.10-3.88)	2.33 (1.60-3.41)	2.81 (2.08-3.79)
	Diagnosed 1-4.9 years ago	2.46 (2.06-2.95)	3.49 (2.61-4.68)	2.40 (1.99-2.90)	2.53 (2.05-3.11)	2.48 (2.07-2.97)
	Diagnosed ≥5 years ago	1.61 (1.39-1.87)	1.45 (1.06-1.97)	1.61 (1.38-1.89)	1.55 (1.30-1.85)	1.63 (1.40-1.89)
Reduced kidney function[d]	Estimated GFR 30-60	1.33 (1.28-1.40)	1.49 (1.36-1.63)	1.33 (1.27-1.39)	1.37 (1.30-1.44)	1.33 (1.27-1.39)
	Estimated GFR <30	2.52 (2.33-2.72)	2.98 (2.57-3.45)	2.47 (2.28-2.68)	2.50 (2.29-2.74)	2.50 (2.31-2.70)
Liver disease		1.75 (1.51-2.03)	1.92 (1.48-2.49)	1.69 (1.44-1.97)	1.75 (1.48-2.07)	1.75 (1.51-2.03)
Stroke/dementia		2.16 (2.06-2.27)	1.74 (1.58-1.93)	2.12 (2.01-2.22)	2.16 (2.05-2.28)	2.16 (2.06-2.27)
Other neurological disease		2.58 (2.38-2.79)	2.26 (1.91-2.68)	2.50 (2.30-2.73)	2.53 (2.31-2.77)	2.58 (2.38-2.80)
Organ transplant		3.53 (2.77-4.49)	2.55 (1.59-4.10)	3.70 (2.89-4.73)	3.45 (2.62-4.54)	3.48 (2.74-4.44)
Asplenia		1.34 (0.98-1.83)	1.87 (1.12-3.11)	1.29 (0.93-1.80)	1.34 (0.94-1.92)	1.33 (0.98-1.82)
Rheumatoid/Lupus/ Psoriasis		1.19 (1.11-1.27)	1.29 (1.14-1.46)	1.17 (1.09-1.25)	1.15 (1.07-1.24)	1.20 (1.12-1.28)
Other immunosuppressive condition		2.21 (1.68-2.90)	2.60 (1.65-4.09)	2.11 (1.58-2.83)	2.24 (1.66-3.03)	1.67 (1.31-2.11)

Models were adjusted for age using a four-knot cubic spline for age, except for estimation of age-group hazard ratios.

[a] Ethnicity hazard ratios in the primary analysis were estimated from a model restricted to those with recorded ethnicity.

[b] For OCS use, 'recent' refers to <1 year before baseline.

[c] HbA1c classification is based on the most recent measurement in the 15 months prior to baseline.

[d] eGFR is measured in ml min^{-1} per 1.73 m^2 and taken from the most recent serum creatinine measurement.

nature research

Corresponding author(s): Ben Godlacre

Last updated by author(s): Jun 16, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected using TPP SystmOne software (14th May maintenance release), for the purpose of direct clinical care. Data management was performed using Python 3.8 and SQL. All code for data management and analysis is archived online at https://github.com/opensafely/risk-factors-research.
Data analysis	Analysis was carried out using Stata 16.1 / Python 3.8. All code for data management and analysis is archived at https://github.com/opensafely/risk-factors-research.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data were linked, stored and analysed securely within the OpenSAFELY platform https://opensafely.org/. All code is shared openly for review and re-use under MIT open license. Detailed pseudonymised patient data is potentially re-identifiable and therefore not shared. All clinical and medicines codelists are openly available for inspection and reuse at https://codelists.opensafely.org/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We conducted a quantitative cohort study using national primary care electronic health record data linked to COVID-19 death data.
Research sample	We used patient data from general practice (GP) records managed by the GP software provider The Phoenix Partnership (TPP), linked to Office for National Statistics (ONS) death data. The sample of patients represents approximately 40% of the population of England, spread geographically across the whole country.
Sampling strategy	Our study population consisted of all adults (males and females 18 years and above) currently registered as active patients in a TPP general practice in England on 1st February 2020. To be included in the study, participants were required to have at least 1 year of prior follow-up in the GP practice to ensure that baseline patient characteristics could be adequately captured, and to have recorded sex, age, and deprivation (see covariates, below).
Data collection	Data were collected by clinicians (e.g. doctors, nurses) and administrative staff, for the purpose of direct clinical care. This was carried out on computers using TPP SystmOne software. The researchers were not present for data collection into the TPP database. Data were then queried from the TPP database by the researchers, to create the study dataset. This was carried out using Python 3.8 and SQL software (available here https://github.com/opensafely/risk-factors-research). This study did not have an experimental condition or hypothesis.
Timing	Patients were observed from the 1st of February 2020 and were followed until the first of either their death date (whether COVID-19 related or due to other causes) or the study end date, 6th May 2020.
Data exclusions	To be included in the study, participants were required to have at least 1 year of prior follow-up in the GP practice to ensure that baseline patient characteristics could be adequately captured, and to have recorded sex, age, and deprivation. The total number of excluded patients was 6,322,225.
Non-participation	No participants dropped out.
Randomization	Participants were not allocated into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	This study uses data gathered during routine medical practice. We selected all patients except those <18 years old, anyone without a recorded sex, age, or deprivation score, and anyone without a year of prior follow-up (to ensure that baseline patient characteristics could be adequately captured). These inclusive criteria mean that bias is minimised.

Ethics oversight	This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863). In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organisations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure. Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform and set aside the requirement for patient consent for COVID-19 related public health research. GP practices, from which the primary care data is obtained, are required to share relevant health information to support the public health response to the pandemic, and have been informed of the OpenSAFELY analytics platform.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Convergent antibody responses to SARS-CoV-2 in convalescent individuals

https://doi.org/10.1038/s41586-020-2456-9

Received: 3 May 2020

Accepted: 12 June 2020

Davide F. Robbiani[1,9,10]✉, Christian Gaebler[1,10], Frauke Muecksch[2,10], Julio C. C. Lorenzi[1,10], Zijun Wang[1,10], Alice Cho[1,10], Marianna Agudelo[1,10], Christopher O. Barnes[3,10], Anna Gazumyan[1,10], Shlomo Finkin[1,10], Thomas Hägglöf[1,10], Thiago Y. Oliveira[1,10], Charlotte Viant[1,10], Arlene Hurley[4], Hans-Heinrich Hoffmann[5], Katrina G. Millard[1], Rhonda G. Kost[6], Melissa Cipolla[1], Kristie Gordon[1], Filippo Bianchini[1], Spencer T. Chen[1], Victor Ramos[1], Roshni Patel[1], Juan Dizon[1], Irina Shimeliovich[1], Pilar Mendoza[1], Harald Hartweger[1], Lilian Nogueira[1], Maggi Pack[1], Jill Horowitz[1], Fabian Schmidt[2], Yiska Weisblum[2], Eleftherios Michailidis[5], Alison W. Ashbrook[5], Eric Waltari[7], John E. Pak[7], Kathryn E. Huey-Tubman[3], Nicholas Koranda[3], Pauline R. Hoffman[3], Anthony P. West Jr[3], Charles M. Rice[5], Theodora Hatziioannou[2], Pamela J. Bjorkman[3]✉, Paul D. Bieniasz[2,8]✉, Marina Caskey[1]✉ & Michel C. Nussenzweig[1,8]✉

During the coronavirus disease-2019 (COVID-19) pandemic, severe acute respiratory syndrome-related coronavirus-2 (SARS-CoV-2) has led to the infection of millions of people and has claimed hundreds of thousands of lives. The entry of the virus into cells depends on the receptor-binding domain (RBD) of the spike (S) protein of SARS-CoV-2. Although there is currently no vaccine, it is likely that antibodies will be essential for protection. However, little is known about the human antibody response to SARS-CoV-2[1–5]. Here we report on 149 COVID-19-convalescent individuals. Plasma samples collected an average of 39 days after the onset of symptoms had variable half-maximal pseudovirus neutralizing titres; titres were less than 50 in 33% of samples, below 1,000 in 79% of samples and only 1% of samples had titres above 5,000. Antibody sequencing revealed the expansion of clones of RBD-specific memory B cells that expressed closely related antibodies in different individuals. Despite low plasma titres, antibodies to three distinct epitopes on the RBD neutralized the virus with half-maximal inhibitory concentrations (IC_{50} values) as low as 2 ng ml^{-1}. In conclusion, most convalescent plasma samples obtained from individuals who recover from COVID-19 do not contain high levels of neutralizing activity. Nevertheless, rare but recurring RBD-specific antibodies with potent antiviral activity were found in all individuals tested, suggesting that a vaccine designed to elicit such antibodies could be broadly effective.

Between 1 April and 8 May 2020, 157 eligible participants were enrolled in the study. Of these, 111 (70.7%) were individuals who had a SARS-CoV-2 infection, as confirmed by PCR with reverse transcription (RT–PCR) (cases), and 46 (29.3%) were close contacts of individuals diagnosed with SARS-CoV-2 infection (contacts). Although inclusion criteria allowed for enrolment of asymptomatic participants, eight contacts who did not develop symptoms were excluded from further analyses. The 149 cases and contacts were free of symptoms that are suggestive of COVID-19 for at least 14 days at the time of sample collection. Participant demographics and clinical characteristics are shown in Extended Data Fig. 1a and Supplementary Tables 1, 2. Only one individual who tested positive for SARS-CoV-2 infection by RT–PCR remained asymptomatic. The other 148 participants reported symptoms that were suggestive of COVID-19 with a mean time of onset of symptoms of approximately 39 days (range, 17–67 days) before sample collection. In this cohort, symptoms lasted for an average of 12 days (0–35 days), and 11 (7%) of the participants were hospitalized. The most common symptoms were fever (83.9%), fatigue (71.1%), cough (62.4%) and myalgia (61.7%), whereas baseline comorbidities were infrequent (10.7%) (Supplementary Tables 1, 2). There were no significant differences in duration or severity (Methods) of symptoms, or in the time from onset of symptoms to sample collection between genders or between cases and contacts. There was no age difference between women and men in our cohort (Extended Data Fig. 1).

---

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA. [2]Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA. [3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [4]Hospital Program Direction, The Rockefeller University, New York, NY, USA. [5]Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, NY, USA. [6]Center for Clinical Translational Science, The Rockefeller University, New York, NY, USA. [7]Chan Zuckerberg Biohub, San Francisco, CA, USA. [8]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA. [9]Present address: Institute for Research in Biomedicine, Università della Svizzera italiana, Bellinzona, Switzerland. [10]These authors contributed equally: Davide F. Robbiani, Christian Gaebler, Frauke Muecksch, Julio C. C. Lorenzi, Zijun Wang, Alice Cho, Marianna Agudelo, Christopher O. Barnes, Anna Gazumyan, Shlomo Finkin, Thomas Hägglöf, Thiago Y. Oliveira, Charlotte Viant. ✉e-mail: drobbiani@irb.usi.ch; bjorkman@caltech.edu; pbieniasz@rockefeller.edu; mcaskey@rockefeller.edu; nussen@rockefeller.edu

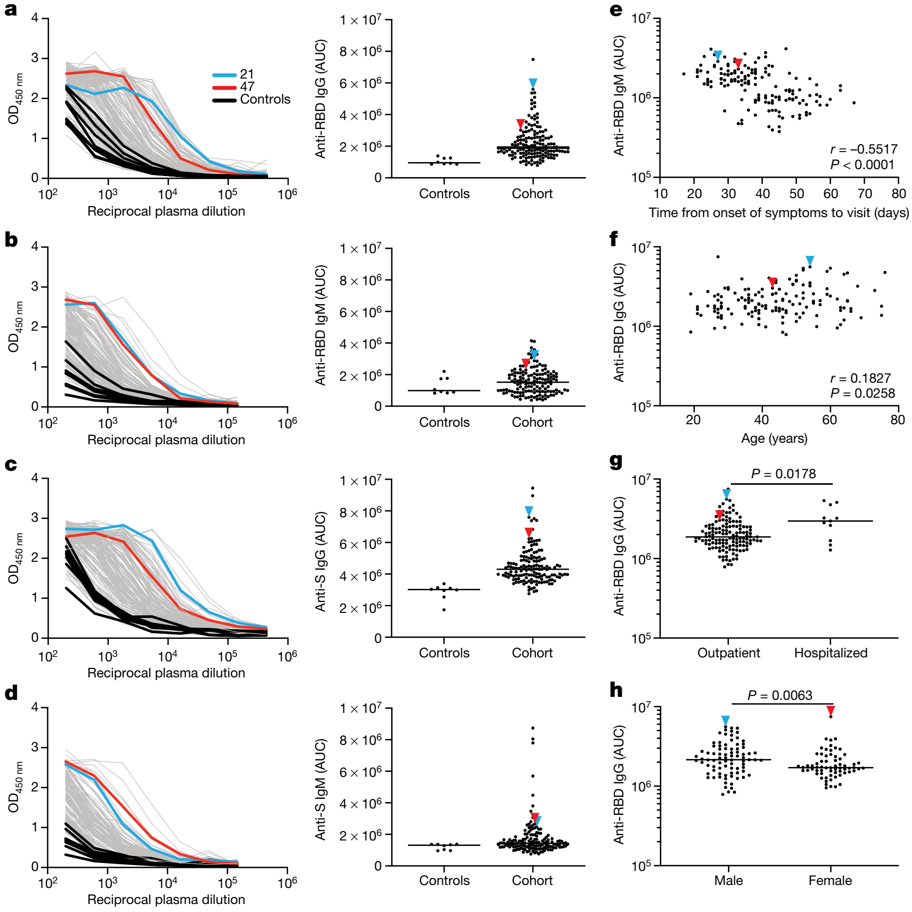


Fig. 1 | Plasma antibodies against SARS-CoV-2. a–d, Results of ELISAs measuring plasma reactivity to RBD (**a**, **b**) and S protein (**c**, **d**). **a**, Anti-RBD IgG. **b**, Anti-RBD IgM. **c**, Anti-S IgG. **d**, Anti-S IgM. Left, optical density at 450 nm ($OD_{450\,nm}$) for the indicated reciprocal plasma dilutions. Right, the normalized area under the curve (AUC) for the 8 controls and 149 individuals in the cohort. Negative controls are shown in black; individuals 21 and 47 are shown as blue and red lines or arrowheads, respectively. **e**, Time between symptom onset and time of sample collection in days is plotted against the normalized AUC for IgM binding to RBD. $r = -0.5517$, $P < 0.0001$. **f**, Participant age in years is plotted against normalized AUC for IgG binding to RBD. $r = 0.1827$ and $P = 0.0258$. The r and P values in **e** and **f** were determined by two-tailed Spearman's correlations. **g**, Normalized AUC of anti-RBD IgG ELISA for outpatients ($n = 138$) and hospitalized individuals ($n = 11$). $P = 0.0178$. **h**, Normalized AUC of anti-RBD IgG ELISA for men ($n = 83$) and women ($n = 66$). $P = 0.0063$. For **g** and **h**, horizontal bars indicate median values. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests.

Plasma samples were tested for binding to the SARS-CoV-2 RBD and trimeric S proteins by a validated enzyme-linked immunosorbent assay (ELISA) using anti-IgG or anti-IgM secondary antibodies for detection[6,7] (Fig. 1, Extended Data Figs. 2, 3 and Supplementary Table 1). Eight independent negative controls and a positive control plasma sample from participant 21 (COV21) were included for normalization of the area under the curve (AUC) in all experiments. Overall, 78% and 70% of the tested plasma samples showed anti-RBD and anti-S IgG AUCs that were at least two standard deviations above the control (Fig. 1a, b). By contrast, only 15% and 34% of the plasma samples showed IgM responses to anti-RBD and anti-S, respectively, that were at least two standard deviations above control (Fig. 1c, d). There was no positive correlation between anti-RBD or anti-S IgG or IgM levels and the duration of symptoms or the timing of sample collection relative to the onset of symptoms (Fig. 1e and Extended Data Fig. 3a–c, g–j). By contrast, as might be expected, anti-RBD IgM titres were negatively correlated with the duration of symptoms and the timing of sample collection (Fig. 1e and Extended Data Fig. 3h). Anti-RBD IgG levels were moderately correlated with age and the severity of symptoms including hospitalization (Fig. 1f, g and Extended Data Fig. 3k). Notably, women had lower anti-RBD and anti-S IgG titres than men (Fig. 1h and Extended Data Fig. 2f).

To measure the neutralizing activity in convalescent plasma samples, we used HIV-1-based virions that carried a nanoluc luciferase reporter, which were pseudotyped with the SARS-CoV-2 S protein (SARS-CoV-2 pseudovirus; Fig. 2, Methods and Extended Data Fig. 4). Negative (historical samples) and positive (COV21) controls were included in all experiments. The overall level of neutralizing activity in the cohort, as measured by the half-maximal neutralizing titre (NT_{50}), was generally low; NT_{50} values were less than 50 in 33% of samples and below 1,000 in 79% of samples (Fig. 2a, b). The geometric mean NT_{50} was 121 (arithmetic mean = 714), and only 2 individuals reached NT_{50} values above 5,000 (Fig. 2a, b and Supplementary Table 1).

Notably, levels of anti-RBD and anti-S IgG antibodies correlated strongly with NT_{50} values (Fig. 2c, d). Neutralizing activity also correlated with age, the duration of symptoms and the severity of symptoms (Extended Data Fig. 5). Consistent with this observation, hospitalized individuals with a longer duration of symptoms showed slightly higher average levels of neutralizing activity than individuals who were not hospitalized ($P = 0.0495$) (Fig. 2e). Finally, we observed a significant difference in neutralizing activity between men and women ($P = 0.0031$) (Fig. 2f). The difference between men and women was consistent with higher anti-RBD and anti-S IgG titres in men, and could not be attributed to age, severity of symptoms, timing of sample collection relative to the onset or duration of symptoms (Fig. 1h and Extended Data Figs. 1b–e, 2f).

To determine the nature of the antibodies elicited by SARS-CoV-2 infection, we used flow cytometry to isolate individual B lymphocytes that carried receptors that bound to the RBD from the blood

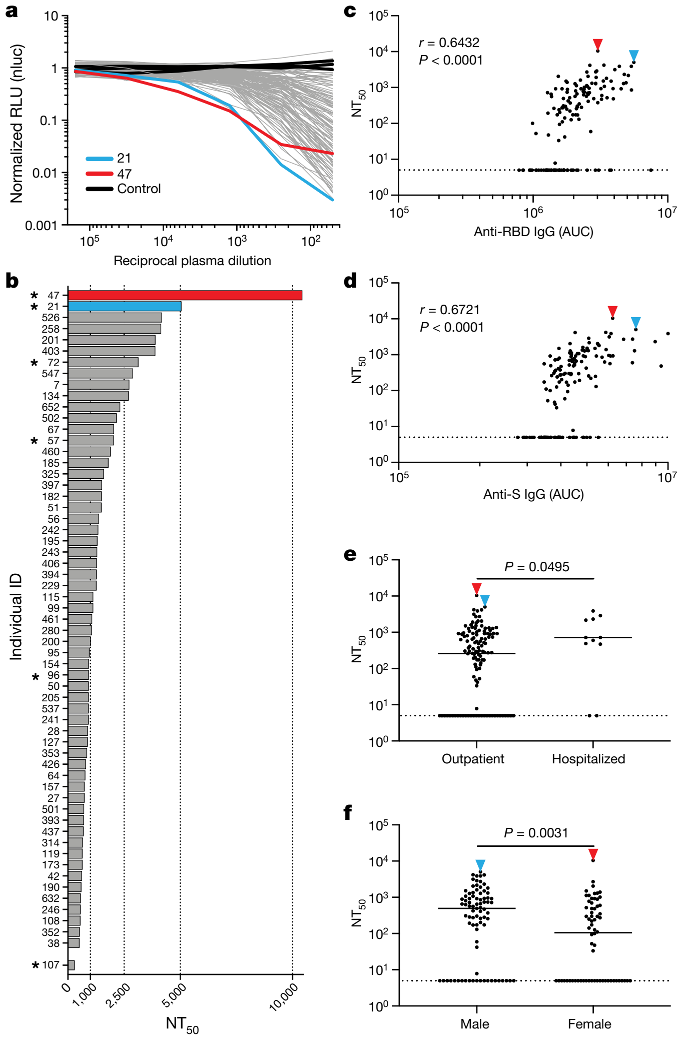


Fig. 2 | Neutralization of SARS-CoV-2 pseudovirus by plasma. **a**, The normalized relative luminescence values (RLU) for cell lysates of $293T_{ACE2}$ cells 48 h after infection with nanoluc-expressing SARS-CoV-2 pseudovirus in the presence of increasing concentrations of plasma derived from 149 participants (grey; except individuals 21 and 47, for which data are shown in blue and red, respectively) and 3 negative controls (black lines). Data are the mean of duplicates; representative of two independent experiments. **b**, Ranked average half-maximal inhibitory plasma neutralizing titre (NT_{50}) for the 59 out of 149 individuals with $NT_{50} > 500$ and individual 107. Asterisks indicate donors from whom antibody sequences were derived. **c**, Normalized AUC for anti-RBD IgG ELISA plotted against NT_{50} values. $r = 0.6432$, $P < 0.0001$. **d**, Normalized AUC for anti-S IgG ELISA plotted against NT_{50} values. $r = 0.6721$, $P < 0.0001$. The r and P values in **c** and **d** were determined by two-tailed Spearman's correlations. **e**, NT_{50} values for outpatients ($n = 138$) and hospitalized individuals ($n = 11$). $P = 0.0495$. **f**, NT_{50} values for men ($n = 83$) and women ($n = 66$) in the cohort. $P = 0.0031$. Statistical significance in **e** and **f** was determined using two-tailed Mann–Whitney U-tests and horizontal bars indicate median values. Dotted lines in **c**–**f** ($NT_{50} = 5$) represent the lower limit of detection. Samples with neutralizing titres below 50 were plotted at the lower limit of detection.

of six selected individuals, including the two samples with top neutralizing activity and four samples with high-to-intermediate neutralizing activity (Fig. 3). The frequency of antigen-specific B cells, identified by their ability to bind to both phycoerythrin (PE)- and AlexaFluor-647-labelled RBD, ranged from 0.07 to 0.005% of all circulating B cells in COVID-19-convalescent individuals, whereas they were undetectable in pre-COVID-19 control samples (Fig. 3a and Extended Data Fig. 6). We obtained 534 paired IgG heavy and light chain (IGH and IGL) sequences by RT–PCR from individual RBD-binding B cells from the 6 convalescent individuals (Methods and Supplementary Table 3). When compared to the human antibody repertoire, several *IGHV* and *IGLV* genes were significantly overrepresented (Extended Data Fig. 7). The average number of nucleotide mutations in V genes for IGH and IGL was 4.2 and 2.8, respectively (Extended Data Fig. 8), which is lower than in antibodies cloned from individuals with chronic infections such as hepatitis B or HIV-1, and similar to antibodies derived from individuals with a primary malaria infection or from non-antigen-enriched circulating IgG memory cells[8–11]. Among other antibody features, IGH CDR3 length was indistinguishable from the reported norm and hydrophobicity was below average[12] (Extended Data Fig. 8).

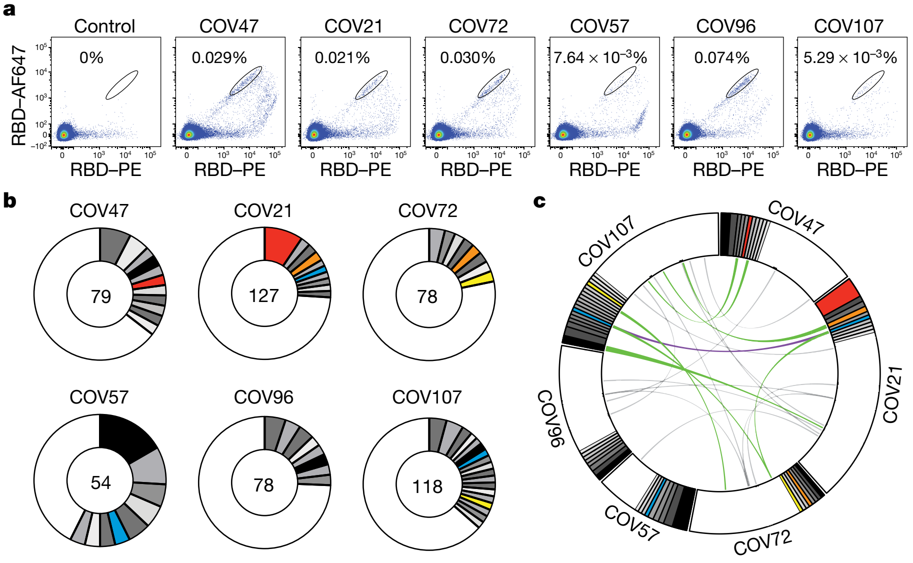
As is the case with other human pathogens, there were expanded clones of viral antigen-binding B cells in all tested individuals convalescent after COVID-19 (Fig. 3b, c and Methods). Overall, 32.2% of the recovered IGH and IGL sequences were from clonally expanded B cells (range, 21.8–57.4% across individuals) (Fig. 3b). Antibodies that shared specific combinations of *IGHV* and *IGLV* genes in different individuals comprised 14% of all the clonal sequences (Fig. 3b, c). Notably, the amino acid sequences of some antibodies found in different individuals were nearly identical (Fig. 3d). For example, antibodies expressed by clonally expanded B cells with IGHV1-58/IGKV3-20 and IGHV3-30-3/IGKV1-39 were found repeatedly in different individuals and had amino acid sequence identities of up to 99% and 92%, respectively (Fig. 3d and Supplementary Table 4). We conclude that the IgG memory response to the SARS-CoV-2 RBD is rich in recurrent and clonally expanded antibody sequences.

To examine the binding properties of anti-SARS-CoV-2 antibodies, we expressed 94 representative antibodies, 67 from clones and 27 from singlets (Supplementary Table 5). ELISAs showed that 95% (89 out of 94) of the antibodies tested including clonal and unique sequences bound to the SARS-CoV-2 RBD with an average half-maximal effective concentration (EC_{50}) of 6.9 ng ml^{-1} (Fig. 4a and Extended Data Fig. 9a). A fraction of these (7 out of 77 that were tested, or 9%) cross-reacted with the RBD of SARS-CoV with EC_{50} values below 1 μg ml^{-1} (Extended Data Fig. 9b, c). No significant cross-reactivity was noted to the RBDs of MERS, HCoV-OC43, HCoV-229E or HCoV-NL63.

To determine whether the monoclonal antibodies had neutralizing activity, we tested them against the SARS-CoV-2 pseudovirus (Fig. 4 and Supplementary Table 6). Among 89 RBD-binding antibodies tested, we found 52 that neutralized SARS-CoV-2 pseudovirus with IC_{50} values ranging from 3 to 709 ng ml^{-1} (Fig. 4b, c, e and Supplementary Table 6). A subset of the most potent of these antibodies was also tested against authentic SARS-CoV-2 and these antibodies neutralized the virus with IC_{50} values of less than 5 ng ml^{-1} (Fig. 4d, e). Only two of the antibodies that cross-reacted with the RBD of SARS-CoV showed significant neutralizing activity against SARS-CoV pseudovirus (Extended Data Fig. 9d, e).

Potent neutralizing antibodies were found in individuals irrespective of their plasma NT_{50} values. For example, antibodies C121, C144 and C135, which had IC_{50} values of 1.64, 2.55 and 2.98 ng ml^{-1} against authentic SARS-CoV-2, respectively, were obtained from individuals COV107, COV47 and COV72, for whom the plasma NT_{50} values were 297, 10,433 and 3,138, respectively (Figs. 2b, 4). Finally, antibodies with recurrent combinations of *IGHV* and *IGLV* genes were among the strongest neutralizing antibodies—for example, antibody C002 is composed of IGHV3-30/IGKV1-39 and shared by the two donors with the strongest plasma neutralizing activity (Figs. 3b, 4). We conclude that even individuals with modest plasma neutralizing activity have rare IgG memory B cells that produce potent SARS-CoV-2-neutralizing antibodies.

To determine whether human anti-SARS-CoV-2 monoclonal antibodies with neutralizing activity can bind to distinct domains on the RBD, we performed bilayer interferometry experiments in which a preformed antibody–RBD immune complex was exposed to a second monoclonal antibody. The antibodies tested comprised three groups, all of which differed in their binding properties from CR3022, an antibody



d

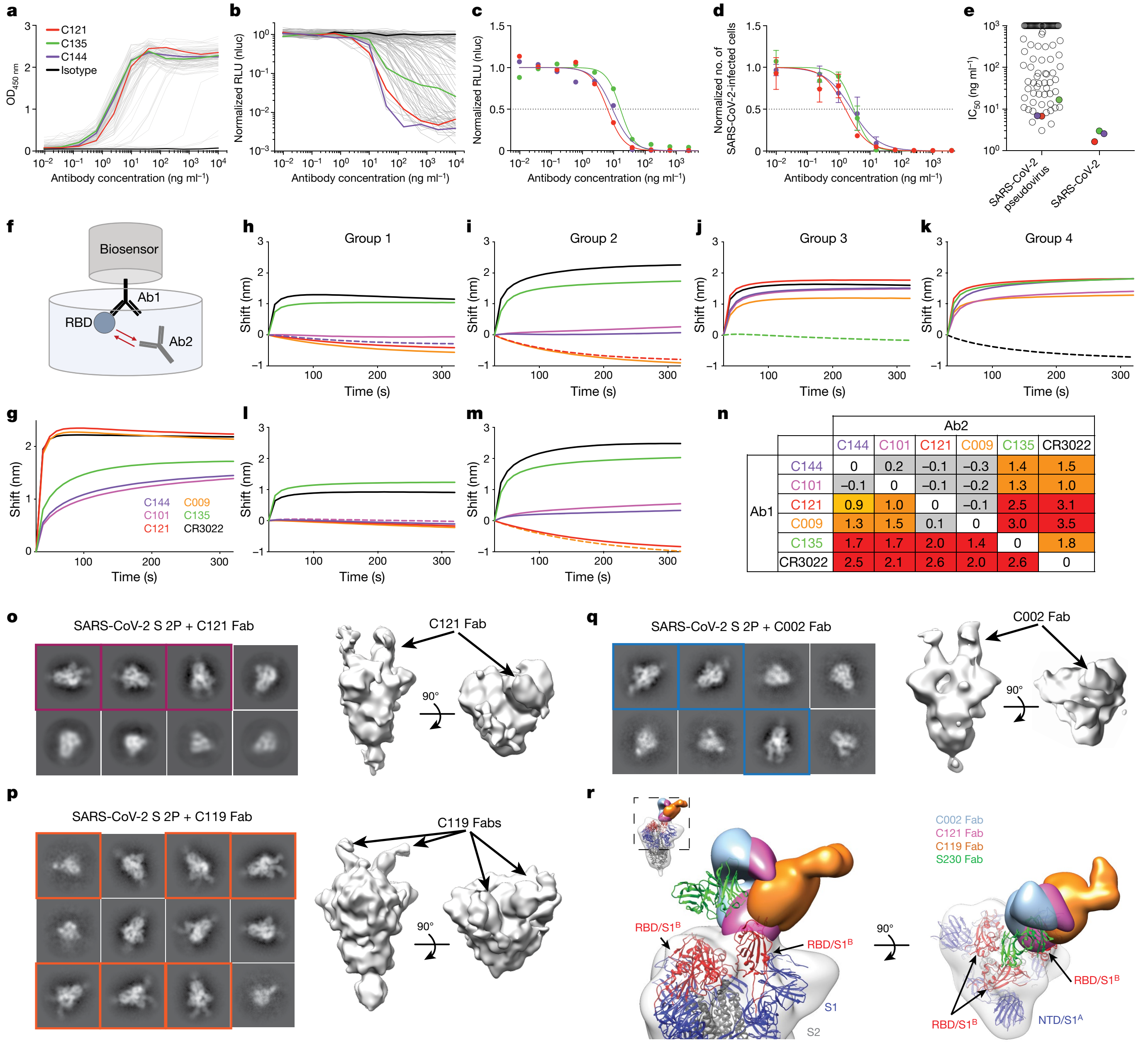
	Heavy				Light		
	IGHV	IGHD	IGHJ	CDRH3	IGLV	IGLJ	CDRL3
COV21	IGHV1-58*01	IGHD2-15*01	IGHJ3*02	**AAPHCSGGSCLDAFDI**	IGKV3-20*01	IGKJ1*01	**QQYGSSPWT**
	IGHV1-58*01	IGHD2-15*01	IGHJ3*02	..........Y.....	IGKV3-20*01	IGKJ1*01	....
COV57	IGHV1-58*02	IGHD2-15*01	IGHJ3*02	..N.......Y.G...	IGKV3-20*01	IGKJ1*01	........M
	IGHV1-58*02	IGHD2-15*01	IGHJ3*02	...Y......N.....	IGKV3-20*01	IGKJ1*01	....
COV107	IGHV1-58*01	IGHD2-2*01	IGHJ3*02	......ST..F.....	IGKV3-20*01	IGKJ1*01	....N....
	IGHV1-58*01	IGHD2-15*01	IGHJ3*02	...Y......S.....	IGKV3-20*01	IGKJ1*01	....
COV21	IGHV3-30-3*01	IGHD5-18*01	IGHJ4*02	**ARDGIVDTAMVTWFDY**	IGKV1-39*01	IGKJ1*01	**QQSYSTPPWT**
	IGHV3-30-3*01	IGHD5-24*01	IGHJ4*02	...--QGMATTY-...	IGKV1-39*01	IGKJ1*01	....N.....
	IGHV3-30-3*01	IGHD5-18*01	IGHJ4*02	.........L......	IGKV1-39*01	IGKJ1*01	....
COV72	IGHV3-30-3*01	IGHD5-18*01	IGHJ5*01	...SD...S.......	IGKV1-39*01	IGKJ1*01	....
	IGHV3-30-3*01	IGHD5-18*01	IGHJ5*01	...SD...........	IGKV1-39*01	IGKJ1*01	....

Fig. 3 | Anti-SARS-CoV-2 RBD antibodies. a, Representative flow cytometry plots showing dual AlexaFluor-647–RBD- and PE–RBD-binding B cells for one control and six study individuals (the gating strategy is shown in Extended Data Fig. 6). Percentages of antigen-specific B cells are indicated. The control is a sample from a healthy individual obtained before COVID-19. **b**, The distribution of antibody sequences from six individuals. The number in the inner circle indicates the number of sequences analysed for the individual denoted above the circle. White indicates sequences isolated only once, and grey or coloured pie slices are proportional to the number of clonally related sequences. Red, blue, orange and yellow pie slices indicate clones that share the same *IGHV* and *IGLV* genes. **c**, Sequences from all six individuals with clonal relationships depicted as in **b**. Interconnecting lines indicate the relationship between antibodies that share V and J gene segment sequences at both IGH and IGL. Purple, green and grey lines connect related clones, clones and singles, and singles to each other, respectively. **d**, Sample sequence alignment for antibodies originating from different individuals that display highly similar IGH V(D)J and IGL VJ sequences including CDR3s. Amino acid differences in CDR3s to the reference sequence (bold) are indicated in red, dashes indicate missing amino acids and dots represent identical amino acids.

that neutralizes SARS-CoV and binds to—but does not neutralize—SARS-CoV-2[13,14]. The antibodies of each of the three groups included: C144 and C101 in group 1; C121 and C009 in group 2; C135 in group 3. All of these antibodies could bind to SARS-CoV-2 RBD that was previously immunocomplexed with CR3022. Groups 1 and 2 also bind to the RBD immunocomplexed with group 3 antibody. Groups 1 and 2 differ in that group 1 can bind to the RBD immunocomplexed with group 2 but not vice versa (Fig. 4f–n). We conclude that similar to SARS-CoV, there are multiple distinct neutralizing epitopes on the RBD of SARS-CoV-2.

To further define the binding characteristics of group-1 and group-2 antibodies, we imaged SARS-CoV-2 S–Fab complexes using negative-stain electron microscopy using C002 (group 1, an IGHV3-30/IGKV1-39 antibody, which is clonally expanded in two donors), C119 and C121 (both in group 2) Fabs (Fig. 4f–r and Extended Data Fig. 10). Consistent with the conformational flexibility of the RBD, two-dimensional class averages showed heterogeneity in both occupancy and orientations of bound Fabs for both groups (Fig. 4o–q). The low resolution of negative-stain electron-microscopy reconstructions precludes detailed binding interpretations; however, the results are consistent with Fabs from both groups being able to recognize 'up' and 'down' states of the RBD, as previously described for some antibodies targeting this epitope[15,16]. The three-dimensional reconstructions are also consistent with competition measurements that indicate that group-1 and group-2 antibodies bind to a RBD epitope that is distinct from the epitope bound by antibody CR3022 (Fig. 4f–n) and with a single-particle cryo-electron microscopy structure of a C105–S complex[17]. In addition, the structures suggest that the antibodies bind to the RBD with different angles of approach; group-1 antibodies have an approach angle that is more similar to the approach angle of the SARS-CoV antibody S230[18] (Fig. 4r).

Human monoclonal antibodies with neutralizing activity against pathogens ranging from viruses to parasites have been obtained from naturally infected individuals by single-cell antibody cloning. Several antibodies have been shown to be effective for the protection and treatment of model organisms and in early-phase clinical studies, but only one antiviral monoclonal antibody is currently in clinical use[19]. Antibodies are relatively expensive and more difficult to produce than small-molecule drugs. However, they differ from drugs in that they can engage the host immune system through their constant domains that



		Ab2					
		C144	C101	C121	C009	C135	CR3022
Ab1	C144	0	0.2	−0.1	−0.3	1.4	1.5
	C101	−0.1	0	−0.1	−0.2	1.3	1.0
	C121	0.9	1.0	0	−0.1	2.5	3.1
	C009	1.3	1.5	0.1	0	3.0	3.5
	C135	1.7	1.7	2.0	1.4	0	1.8
	CR3022	2.5	2.1	2.6	2.0	2.6	0

Fig. 4 | Anti-SARS-CoV-2 RBD antibody reactivity. a, Results of ELISAs measuring monoclonal antibody binding to RBD. *n* = 94 samples and 1 isotype control. In all panels, C121, C135, C144 and isotype control are shown in red, green, purple and black, respectively. **b**, The normalized relative luminescence values for cell lysates of $293T_{ACE2}$ cells 48 h after infection with SARS-CoV-2 pseudovirus in the presence of increasing concentrations of monoclonal antibodies. *n* = 89 samples and 1 isotype control. **c**, SARS-CoV-2 pseudovirus neutralization assay. Normalized relative luminescence values were determined in the presence of a titration of monoclonal antibodies C121, C135 and C144. **d**, SARS-CoV-2 real virus neutralization assay. Normalized number of infected cells (determined by dividing the amount of infection per well by the average of control wells infected in the absence of antibodies) were determined in the presence of a titration of monoclonal antibodies C121, C135 and C144. **a**–**d**, Data are representative of two independent experiments. Data are the mean of duplicates (**b**, **c**) or mean ± s.d. of triplicates (**d**). **e**, IC_{50} values for antibodies assayed in **b** and **d**, the mean value of at least two experiments is shown. Samples with IC_{50} values above 1 μg ml^{-1} were plotted at 1 μg ml^{-1}. *n* = 89 (pseudovirus) and *n* = 3 (virus). **f**, Diagram of the biolayer interferometry experiment. **g**, Binding of C144, C101, C121, C009, C135 and CR3022[13,14] to RBD. **h**–**m**, Second antibody (Ab2) binding to preformed first antibody (Ab1)–RBD complexes. Dotted line denotes when Ab1 and Ab2 are the same, and Ab2 is according to the colour-coding in **g**. **h**, **l**, Group 1 antibodies were tested. C144 (**h**) and C101 (**l**) were used as Ab1. **i**, **m**, Group 2 antibodies were tested. C121 (**i**) and C009 (**m**) were used as Ab1. **j**, A group 3 antibody was tested. C135 was used as Ab1. **k**, A group 4 antibody was tested. CR3022 was used as Ab1. **n**, The shift in nanometres after Ab2 binding to the preformed Ab1–RBD complexes. Values are normalized through the subtraction of the autologous antibody control. Data are representative of two experiments. **o**–**q**, Representative two-dimensional class averages and three-dimensional reconstructed volumes for SARS-CoV-2 S 2P trimers complexed with C002 (**q**), C119 (**p**) and C121 (**o**) Fabs. Two-dimensional class averages with observable Fab density are outlined. **r**, Overlay of S–Fab complexes with fully occupied C002 (blue), C121 (magenta) and C119 (orange) Fabs. The SARS-CoV-2 S model from PDB 6VYB was fit into the density. The SARS-CoV monoclonal antibody S230 (PDB 6NB6) is shown as a reference (green ribbon).

bind to Fc gamma receptors on host immune cells[20]. These interactions can enhance immunity and help to clear the pathogen or infected cells; however, they can also lead to disease enhancement during infections with dengue virus[21] and possibly coronavirus[22]. This problem has impeded the development of dengue virus vaccines, but would not interfere with the clinical use of potent neutralizing antibodies that can be modified to prevent interactions with the Fc gamma receptor and that remain protective against viral pathogens[23].

Antibodies are essential elements of most vaccines and will probably be a crucial component of an effective vaccine against SARS-CoV-2[24–26]. Recurrent antibodies have been observed in other infectious diseases and vaccine responses[11,27–30]. The observation that plasma neutralizing activity is low in most convalescent individuals, but that recurrent anti-SARS-CoV-2 RBD antibodies with potent neutralizing activity can be found in individuals with moderate plasma neutralizing activity suggests that humans are intrinsically capable of generating anti-RBD antibodies that potently neutralize SARS-CoV-2. Thus, vaccines that selectively and efficiently induce antibodies that target the RBD of SARS-CoV-2 may be especially effective.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2456-9.

1. Graham, R. L., Donaldson, E. F. & Baric, R. S. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat. Rev. Microbiol*. **11**, 836–848 (2013).
2. Gralinski, L. E. & Baric, R. S. Molecular pathology of emerging coronavirus infections. *J. Pathol*. **235**, 185–195 (2015).
3. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
4. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292 (2020).
5. Jiang, S., Hillyer, C. & Du, L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol*. **41**, 355–359 (2020).
6. Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med*. https://doi.org/10.1038/s41591-020-0913-5 (2020).
7. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2020).
8. Scheid, J. F. et al. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636–640 (2009).
9. Tiller, T. et al. Autoreactivity in human IgG$^+$ memory B cells. *Immunity* **26**, 205–213 (2007).
10. Murugan, R. et al. Clonal selection drives protective memory B cell responses in controlled human malaria infection. *Sci. Immunol*. **3**, eaap8029 (2018).
11. Wang, Q. et al. A combination of human broadly neutralizing antibodies against hepatitis B virus HBsAg with distinct epitopes suppresses escape mutations. *Cell Host Microbe* https://doi.org/10.1016/j.chom.2020.05.010 (2020).
12. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).
13. ter Meulen, J. et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med*. **3**, e237 (2006).
14. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
15. Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039 (2019).
16. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
17. Barnes, C. O. et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* https://doi.org/10.1016/j.cell.2020.06.025 (2020).
18. Zhu, Z. et al. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **104**, 12123–12128 (2007).
19. Salazar, G., Zhang, N., Fu, T. M. & An, Z. Antibody therapies for the prevention and treatment of viral infections. *NPJ Vaccines* **2**, 19 (2017).
20. Bournazos, S. & Ravetch, J. V. Anti-retroviral antibody FcγR-mediated effector functions. *Immunol. Rev*. **275**, 285–295 (2017).
21. Feinberg, M. B. & Ahmed, R. Advancing dengue vaccine development. *Science* **358**, 865–866 (2017).
22. Iwasaki, A. & Yang, Y. The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev. Immunol*. **20**, 339–341 (2020).
23. Van Rompay, K. K. A. et al. A combination of two human monoclonal antibodies limits fetal damage by Zika virus in macaques. *Proc. Natl Acad. Sci. USA* **117**, 7981–7989 (2020).
24. Plotkin, S. A. Correlates of protection induced by vaccination. *Clin. Vaccine Immunol*. **17**, 1055–1065 (2010).
25. Yu, J. et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* eabc6284 (2020).
26. Chandrashekar, A. et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* eabc4776 (2020).
27. Scheid, J. F. et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
28. Robbiani, D. F. et al. Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico. *Cell* **169**, 597–609 (2017).
29. Ehrhardt, S. A. et al. Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. *Nat. Med*. **25**, 1589–1600 (2019).
30. Pappas, L. et al. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* **516**, 418–422 (2014).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Study participants

Study participants were recruited at the Rockefeller University Hospital in New York from 1 April to 8 May 2020. Eligible participants were adults aged 18–76 years who were either diagnosed with a SARS-CoV-2 infection by RT–PCR and were free of symptoms of COVID-19 for at least 14 days (cases), or who were close contacts (for example, household members, co-workers or members of same religious community) of someone who had been diagnosed with a SARS-CoV-2 infection by RT–PCR and were free of symptoms suggestive of COVID-19 for at least 14 days (contacts). Exclusion criteria included the presence of symptoms suggestive of an active SARS-CoV-2 infection, or haemoglobin levels of <12 g/dl for men and <11 g/dl for women.

Most study participants were residents of the Greater New York City tristate region and were enrolled sequentially according to eligibility criteria. Participants were first interviewed by phone to collect information on their clinical presentation, and subsequently presented to the Rockefeller University Hospital for the collection of a single blood sample. Participants were asked to rate the highest severity of their symptoms on a numeric rating scale ranging from 0 to 10. The score was adapted from the pain scale chart, in which 0 was the lack of symptoms, 4 was distressing symptoms (for example, fatigue, myalgia, fever, cough or shortness of breath) that interfered with daily living activities, 7 was disabling symptoms that prevented the performance of daily living activities, and 10 was unimaginable/unspeakable discomfort (in this case, distress owing to shortness of breath). All participants provided written informed consent before participation in the study and the study was conducted in accordance with Good Clinical Practice and clinical data collection and management were carried out using the software iRIS by iMedRIS. The study was performed in compliance with all relevant ethical regulations and the protocol for studies with human participants was approved by the Institutional Review Board of the Rockefeller University.

Blood samples processing and storage

Peripheral blood mononuclear cells were obtained by gradient centrifugation and stored in liquid nitrogen in the presence of fetal calf serum (FCS) and DMSO. Heparinized plasma and serum samples were aliquoted and stored at −20 °C or less. Before experiments, aliquots of plasma samples were heat-inactivated (56 °C for 1 h) and then stored at 4 °C.

Cloning, expression and purification of recombinant coronavirus proteins

Codon-optimized nucleotide sequences encoding the SARS-CoV-2 S ectodomain (residues 16–1206) and RBD (residues 331–524) were synthesized and subcloned into the mammalian expression pTwist-CMV BetaGlobin vector by Twist Bioscience Technologies based on an early SARS-CoV-2 sequence isolate (GenBank MN985325.1). The SARS-CoV-2 RBD construct included an N-terminal human IL-2 signal peptide and dual C-terminal tags ($(GGGGS)_2$-HHHHHHHH (octa-histidine) and GLNDIFEAQKIEWHE (AviTag)). In addition, the corresponding $S1^B$ or RBDs for SARS-CoV (residues 318–510; GenBank AAP13441.1), MERS-CoV (residues 367–588; GenBank JX869059.2), HCoV-NL63 (residues 481–614; GenBank AAS58177.1), HCoV-OC43 (residues 324–632; GenBank AAT84362.1) and HCoV-229E (residues 286–434; GenBank AAK32191.1) were synthesized with the same N- and C-terminal extensions as the SARS-CoV-2 RBD construct and subcloned into the mammalian expression pTwist-CMV BetaGlobin vector (Twist Bioscience Technologies). The SARS-CoV-2 S ectodomain was modified as previously described[4]. In brief, the S ectodomain construct included an N-terminal mu-phosphatase signal peptide, 2P stabilizing mutations (K986P and V987P), mutations to remove the S1/S2 furin cleavage site (${}_{682}RRAR_{685}$ to GSAS), a C-terminal extension (IKGSG-RENLYFQG (TEV protease site), GGGSG-YIPEAPRDGQAYVRKDGEWVLLSTFL (foldon trimerization motif), G-HHHHHHHH (octa-histidine tag) and GLNDIFEAQKIEWHE (AviTag)). The SARS-CoV-2 S 2P ectodomain and RBD constructs were produced by transient transfection of 500 ml of Expi293F cells (Thermo Fisher) and purified from clarified transfected cell supernatants 4 days after transfection using Ni^{2+}-NTA affinity chromatography (GE Life Sciences). Affinity-purified proteins were concentrated and further purified by size-exclusion chromatography using a Superdex200 16/60 column (GE Life Sciences) running in 1× TBS (20 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.02% NaN_3). Peak fractions were analysed by SDS–PAGE, and fractions corresponding to soluble S 2P trimers or monomeric RBD proteins were pooled and stored at 4 °C.

ELISAs

Validated ELISAs[6,7] to evaluate antibodies binding to SARS-CoV-2 RBD and trimeric spike proteins, and to SARS-CoV RBD, were performed by coating of high-binding 96-half-well plates (Corning 3690) with 50 μl per well of a 1μg/ml protein solution in PBS overnight at 4 °C. Plates were washed 6 times with washing buffer (1× PBS with 0.05% Tween-20 (Sigma-Aldrich)) and incubated with 170 μl per well blocking buffer (1× PBS with 2% BSA and 0.05% Tween-20 (Sigma)) for 1 h at room temperature. Immediately after blocking, monoclonal antibodies or plasma samples were added in PBS and incubated for 1 h at room temperature. Plasma samples were assayed at a 1:200 starting dilution and 7 additional threefold serial dilutions. Monoclonal antibodies were tested at 10 μg/ml starting concentration and 10 additional fourfold serial dilutions. Plates were washed 6 times with washing buffer and then incubated with anti-human IgG or IgM secondary antibody conjugated to horseradish peroxidase (HRP) (Jackson Immuno Research 109-036-088 and 109-035-129) in blocking buffer at a 1:5,000 dilution. Plates were developed by addition of the HRP substrate, TMB (ThermoFisher) for 10 min, then the developing reaction was stopped by adding 50 μl 1 M H_2SO_4 and absorbance was measured at 450 nm with an ELISA microplate reader (FluoStar Omega, BMG Labtech) with Omega and Omega MARS software for analysis. For plasma samples, a positive control (plasma from patient COV21, diluted 200-fold in PBS) and negative control (historical plasma samples) samples were added in duplicate to every assay plate for validation. The average of its signal was used for normalization of all of the other values on the same plate with Excel software before calculating the area under the curve using Prism 8 (GraphPad). For monoclonal antibodies, the EC_{50} was determined using four-parameter nonlinear regression (GraphPad Prism).

$293T_{ACE2}$ cells

For constitutive expression of ACE2 in 293T cells, a cDNA encoding ACE2, carrying two inactivating mutations in the catalytic site (H374N and H378N), was inserted into CSIB 3′ to the SFFV promoter[31]. $293T_{ACE2}$ cells were generated by transduction with CSIB-based virus followed by selection with 5 μg/ml blasticidin.

SARS-CoV-2 and SARS-CoV pseudotyped reporter viruses

A plasmid expressing a C-terminally truncated SARS-CoV-2 S protein (pSARS-CoV2-S_{trunc}) was generated by insertion of a human codon-optimized cDNA encoding SARS-CoV-2 S lacking the C-terminal 19 codons (Geneart) into pCR3.1. The S open-reading frame was taken from 'Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1' (GenBank: NC_045512). For expression of the full-length SARS-CoV S protein, 'human SARS coronavirus spike glycoprotein gene ORF cDNA clone expression plasmid (codon optimized)' (here referred to as pSARS-CoV-S) was obtained from SinoBiological (VG40150-G-N).

An *env*-inactivated HIV-1 reporter construct (pNL4-3ΔEnv-nanoluc) was generated from pNL4-3[32] by introducing a 940-bp deletion 3′ in the *vpu* stop codon, resulting in a frameshift in *env*. The human codon-optimized nanoluc Luciferase reporter gene (Nluc, Promega) was inserted in place of nucleotides 1–100 of the *nef* gene. To generate pseudotyped viral stocks, 293T cells were transfected with pNL4-3ΔEnv-nanoluc and pSARS-CoV2-S_{trunc} or pSARS-CoV-S using polyethylenimine. Co-transfection of pNL4-3ΔEnv-nanoluc and S-expression plasmids leads to production of HIV-1-based virions that carried either the SARS-CoV-2 or SARS-CoV S protein on the surface. After transfection for 8 h, cells were washed twice with PBS and fresh medium was added. Supernatants containing virions were collected 48 h after transfection, filtered and stored at −80 °C. Infectivity of virions was determined by titration on $293T_{ACE2}$ cells. Further details are described elsewhere[33].

Pseudotyped virus neutralization assay

Fivefold serially diluted plasma from COVID-19-convalescent individuals and healthy donors or fourfold serially diluted monoclonal antibodies were incubated with the SARS-CoV-2 or SARS-CoV pseudotyped virus for 1 h at 37 °C. The mixture was subsequently incubated with $293T_{ACE2}$ cells for 48 h after which cells were washed twice with PBS and lysed with Luciferase Cell Culture Lysis 5× reagent (Promega). Nanoluc Luciferase activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega) with Modulus II Microplate Reader User interface (TURNER BioSystems). The obtained relative luminescence units were normalized to those derived from cells infected with SARS-CoV-2 or SARS-CoV pseudotyped virus in the absence of plasma or monoclonal antibodies. The half-maximal inhibitory concentration for plasma (NT_{50}) or monoclonal antibodies (IC_{50}) was determined using four-parameter nonlinear regression (GraphPad Prism).

Cell lines, virus and virus titration

VeroE6 kidney epithelial cells (*Chlorocebus sabaeus*; ATCC) and Huh-7.5 hepatoma cells (*Homo sapiens*; C.M.R.) were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 1% nonessential amino acids and 10% FCS at 37 °C and 5% CO_2. All cell lines have been tested negative for contamination with mycoplasma and were obtained from the ATCC (with the exception for Huh-7.5). SARS-CoV-2, strain USA-WA1/2020, was obtained from BEI Resources and amplified in VeroE6 cells at 33 °C. Viral titres were measured on Huh-7.5 cells by standard plaque assay. In brief, 500 μl of serial tenfold virus dilutions in Opti-MEM were used to infect 400,000 cells seeded the previous day in a 6-well plate format. After 90 min adsorption, the virus inoculum was removed, and cells were overlayed with DMEM containing 10% FCS with 1.2% microcrystalline cellulose (Avicel). Cells were incubated for 5 days at 33 °C, followed by fixation with 3.5% formaldehyde and crystal violet staining for plaque enumeration. All experiments were performed in a biosafety level 3 laboratory.

Microscopy-based neutralization assay of authentic SARS-CoV-2

The day before infection, VeroE6 cells were seeded at 12,500 cells/well into 96-well plates. Antibodies were serially diluted in BA-1, mixed with a constant amount of SARS-CoV-2 (grown in VeroE6) and incubated for 60 min at 37 °C. The antibody–virus mix was then directly applied to VeroE6 cells (MOI of ~0.1 PFU/cell). Cells were fixed 18 h after infection by adding an equal volume of 7% formaldehyde to the wells, followed by permeabilization with 0.1% Triton X-100 for 10 min. After extensive washing, cells were incubated for 1 h at room temperature with blocking solution of 5% goat serum in PBS (005–000-121; Jackson ImmunoResearch). A rabbit polyclonal anti-SARS-CoV-2 nucleocapsid antibody (GTX135357; GeneTex) was added to the cells at 1:500 dilution in blocking solution and incubated at 4 °C overnight. A goat anti-rabbit AlexaFluor 594 (A-11012; Life Technologies) at a dilution of 1:2,000 was used as a secondary antibody. Nuclei were stained with Hoechst 33342 (62249; Thermo Scientific) at a 1:1,000 dilution. Images were acquired with a fluorescence microscope and analysed using ImageXpress Micro XLS and MetaXpress software (Molecular Devices). All statistical analyses were done using Prism 8 software (GraphPad).

Biotinylation of viral protein for use in flow cytometry

Purified and Avi-tagged SARS-CoV-2 RBD was biotinylated using the Biotin-Protein Ligase-BIRA kit according to manufacturer's instructions (Avidity). Ovalbumin (Sigma, A5503-1G) was biotinylated using the EZ-Link Sulfo-NHS-LC-Biotinylation kit according to the manufacturer's instructions (Thermo Scientific). Biotinylated ovalbumin was conjugated to streptavidin-BV711 (BD biosciences, 563262) and RBD to streptavidin-PE (BD Biosciences, 554061) and streptavidin-AF647 (Biolegend, 405237)[34].

Single-cell sorting by flow cytometry

Peripheral blood mononuclear cells were enriched for B cells by negative selection using a pan-B-cell isolation kit according to the manufacturer's instructions (Miltenyi Biotec, 130-101-638). The enriched B cells were incubated in FACS buffer (1× PBS, 2% FCS, 1 mM EDTA) with the following anti-human antibodies (all at 1:200 dilution): anti-CD20-PECy7 (BD Biosciences, 335793), anti-CD3-APC-eFluro 780 (Invitrogen, 47-0037-41), anti-CD8-APC-eFluor 780 (Invitrogen, 47-0086-42), anti-CD16-APC-eFluor 780 (Invitrogen, 47-0168-41), anti-CD14-APC-eFluor 780 (Invitrogen, 47-0149-42), as well as Zombie NIR (BioLegend, 423105) and fluorophore-labelled RBD and ovalbumin (Ova) for 30 min on ice[34]. Single $CD3^-CD8^-CD14^-CD16^-CD20^+Ova^-$RBD-$PE^+$RBD-$AF647^+$ B cells were sorted into individual wells of 96-well plates containing 4 μl of lysis buffer (0.5× PBS, 10 mM DTT, 3,000 units/ml RNasin Ribonuclease Inhibitors (Promega, N2615) per well using a FACS Aria III and FACSDiva software (Becton Dickinson) for acquisition and FlowJo for analysis. The sorted cells were frozen on dry ice, and then stored at −80 °C or immediately used for subsequent RNA reverse transcription. Although cells were not stained for IgG expression, they are memory B cells based on the fact that they are $CD20^+$ (a marker that is absent in plasmablasts) and they express IgG (as antibodies were amplified from these cells using IgG-specific primers).

Antibody sequencing, cloning and expression

Antibodies were identified and sequenced as described previously[28,35,36]. In brief, RNA from single cells was reverse-transcribed (SuperScript III Reverse Transcriptase, Invitrogen, 18080-044) and the cDNA stored at −20 °C or used for subsequent amplification of the variable IGH, IGL and IGK genes by nested PCR and Sanger sequencing[35]. Anti-Zika virus monoclonal antibody Z021[28] was used as isotype control. Sequence analysis was performed using MacVector. Amplicons from the first PCR reaction were used as templates for sequence- and ligation-independent cloning into antibody expression vectors. Recombinant monoclonal antibodies and Fabs were produced and purified as previously described[37,38].

Biolayer interferometry

Biolayer interferometry assays were performed on the Octet Red instrument (ForteBio) at 30 °C with shaking at 1,000 r.p.m. Epitope-binding assays were performed with protein A biosensor (ForteBio 18-5010), following the manufacturer's protocol 'classical sandwich assay'. (1) Sensor check: sensors immersed 30 s in buffer alone (buffer ForteBio 18-1105). (2) Capture first antibody: sensors immersed 10 min with Ab1 at 40 μg/ml. (3) Baseline: sensors immersed 30 s in buffer alone. (4) Blocking: sensors immersed 5 min with IgG isotype control at 50 μg/ml. (6) Antigen association: sensors immersed 5 min with RBD at 100 μg/ml. (7) Baseline: sensors immersed 30 s in buffer alone. (8) Association Ab2: sensors immersed 5 min with Ab2 at 40 μg/ml. Curve fitting was performed using the Fortebio Octet Data analysis software (ForteBio).

Computational analyses of antibody sequences

Antibody sequences were trimmed based on quality and annotated using Igblastn v.1.14.0[39] with IMGT domain delineation system. Annotation was performed systematically using Change-O toolkit v.0.4.5[40]. Heavy and light chains derived from the same cell were paired, and clonotypes were assigned based on their V and J genes using in-house R and Perl scripts (Fig. 3b, c). All scripts and the data used to process antibody sequences are publicly available on GitHub (https://github.com/stratust/igpipeline).

The frequency distributions of human V genes in anti-SARS-CoV-2 antibodies from this study were compared to Sequence Read Archive accession SRP010970[41]. The V(D)J assignments were done using IMGT/High V-Quest and the frequencies of heavy and light chain V genes were calculated for 14 and 13 individuals, respectively, using sequences with unique CDR3s. The two-tailed *t*-test with unequal variances was used to determine statistical significance (Extended Data Fig. 7).

Nucleotide somatic hypermutation and CDR3 length were determined using in-house R and Perl scripts. For somatic hypermutations, IGHV and IGLV nucleotide sequences were aligned against their closest germlines using Igblastn and the number of differences were considered nucleotide mutations. The average mutations for V genes was calculated by dividing the sum of all nucleotide mutations across all patients by the number of sequences used for the analysis. To calculate the GRAVY scores of hydrophobicity[42] we used Guy H.R. Hydrophobicity scale based on free energy of transfer (kcal/mole)[43] implemented by the R package Peptides (the Comprehensive R Archive Network repository; https://journal.r-project.org/archive/2015/RJ-2015-001/RJ-2015-001.pdf). We used 533 heavy chain CDR3 amino acid sequences from this study (sequence COV047_P4_IgG_51-P1369 lacks CDR3 amino acid sequence) and 22,654,256 IGH CDR3 sequences from the public database of memory B cell receptor sequences[44]. The Shapiro–Wilk test was used to determine whether the GRAVY scores are normally distributed. The GRAVY scores from all 533 IGH CDR3 amino acid sequences from this study were used to perform the test and 5,000 GRAVY scores of the sequences from the public database were randomly selected. The Shapiro–Wilk *P* values were 6.896×10^{-3} and 2.217×10^{-6} for sequences from this study and the public database, respectively, indicating that the data were not normally distributed. Therefore, we used the Wilcoxon nonparametric test to compare the samples, which indicated a difference in hydrophobicity distribution ($P = 5 \times 10^{-6}$) (Extended Data Fig. 8).

Negative-stain electron-microscopy data collection and processing

Purified Fabs (C002, C119 and C121) were complexed with SARS-CoV-2 S trimer at a twofold molar excess for 1 min and diluted to 40 μg/ml in TBS immediately before adding 3 μl to a freshly glow-discharged ultrathin, 400-mesh carbon-coated copper grid (Ted Pella). Samples were blotted after a 1-min incubation period and stained with 1% uranyl formate for an additional minute before imaging. Micrographs were recorded on a Thermo Fisher Talos Arctica transmission electron microscope operating at 200 keV using a K3 direct electron detector (Gatan) and SerialEM automated image-acquisition software[45]. Images were acquired at a nominal magnification of 28,000× (1.44 Å/pixel size) and a defocus range of −1.5 to −2.0 μm. Images were processed in cryoSPARC, and reference-free particle picking was completed using a Gaussian blob picker[46]. Reference-free two-dimensional class averages and ab initio volumes were generated in cryoSPARC, and subsequently three-dimensionally classified to identify classes of S–Fab complexes, that were then homogenously refined. Figures were prepared using UCSF Chimera[47]. The resolutions of the final single-particle reconstructions were about 17–20 Å calculated using a gold-standard FSC (0.143 cut-off) and about 24–28 Å using a 0.5 cut-off.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data are provided in Supplementary Tables 1, 3–6. The raw sequencing data associated with Fig. 3 has been deposited at Github (https://github.com/stratust/igpipeline). This study also uses data from 'A public database of memory and naive B-cell receptor sequences' (https://doi.org/10.5061/dryad.35ks2), from PDB (6VYB and 6NB6) and from NCBI Sequence Read Archive (SRP010970).

Code availability

Computer code to process the antibody sequences is available at GitHub (https://github.com/stratust/igpipeline).

31. Kane, M. et al. Identification of interferon-stimulated genes with antiretroviral activity. *Cell Host Microbe* **20**, 392–405 (2016).
32. Adachi, A. et al. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol*. **59**, 284–291 (1986).
33. Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med*. https://doi.org/10.1084/jem.20201181 (2020).
34. Wang, Z. et al. Isolation of single HIV-1 envelope specific B cells and antibody cloning from immunized rhesus macaques. *J. Immunol. Methods* **478**, 112734 (2020).
35. Tiller, T. et al. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* **329**, 112–124 (2008).
36. von Boehmer, L. et al. Sequencing and cloning of antigen-specific antibodies from mouse memory B cells. *Nat. Protocols* **11**, 1908–1923 (2016).
37. Klein, F. et al. Enhanced HIV-1 immunotherapy by commonly arising antibodies that target virus escape variants. *J. Exp. Med*. **211**, 2361–2372 (2014).
38. Schoofs, T. et al. Broad and potent neutralizing antibodies recognize the silent face of the HIV envelope. *Immunity* **50**, 1513–1529 (2019).
39. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*. **41**, W34–W40 (2013).
40. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).
41. Rubelt, F. et al. Onset of immune senescence defined by unbiased pyrosequencing of human immunoglobulin mRNA repertoires. *PLoS ONE* **7**, e49774 (2012).
42. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol*. **157**, 105–132 (1982).
43. Guy, H. R. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys. J*. **47**, 61–70 (1985).
44. DeWitt, W. S. et al. A public database of memory and naive B-cell receptor sequences. *PLoS ONE* **11**, e0160853 (2016).
45. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol*. **152**, 36–51 (2005).
46. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
47. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol*. **157**, 281–287 (2007).

Acknowledgements We thank all study participants who devoted time to our research; B. Coller and S. Schlesinger, the Rockefeller University Hospital Clinical Research Support Office and nursing staff; J. L. DeRisi for facilitating interactions with the Chan Zuckerberg Biohub; all members of the M.C.N. laboratory for discussions; A. Escolano, G. Breton and B. Reis; M. Jankovic for laboratory support; and J. Vielmetter and the Protein Expression Center in the Beckman Institute at Caltech. This work was supported by NIH grant P01-AI138398-S1 (M.C.N., C.M.R. and P.J.B.) and 2U19AI111825 (M.C.N. and C.M.R.); the Caltech Merkin Institute for Translational Research and P50 AI150464 (P.J.B.), George Mason University Fast Grants (D.F.R. and P.J.B.) and the European ATAC consortium EC 101003650 (D.F.R.); R01-AI091707-10S1 to C.M.R.; R37-AI64003 to P.D.B.; R01AI78788 to T. Hatziioannou; The G. Harold and Leila Y. Mathers Charitable Foundation to C.M.R. Electron microscopy was performed in the Caltech Beckman Institute Resource Center for Transmission Electron Microscopy (directors: S. Chen and A. Malyutin). C.G. was supported by the Robert S. Wennett Post-Doctoral Fellowship, in part by the National Center for Advancing Translational Sciences (National Institutes of Health Clinical and Translational Science Award programme, grant UL1 TR001866) and by the Shapiro-Silverberg Fund for the Advancement of Translational Research. P.D.B. and M.C.N. are investigators of the Howard Hughes Medical Institute.

Author contributions D.F.R., P.D.B., P.J.B., T. Hatziioannou, C.M.R. and M.C.N. conceived, designed and analysed the experiments. D.F.R., M. Caskey and C.G. designed clinical protocols. F.M., J.C.C.L., Z.W., A.C., M.A., C.O.B., S.F., T. Hägglöf, C.V., K.G., F.B., S.T.C., P.M., H.H., L.N., F.S., Y.W., H.-H.H., E.M., A.W.A., K.E.H.-T., N.K. and P.R.H. carried out experiments. A.G. and M. Cipolla

produced antibodies. C.O.B., J.E.P. and E.W. produced SARS-CoV-2 proteins. A.H., R.G.K., J.H., K.G.M., C.G. and M. Caskey recruited participants and executed clinical protocols. R.P., J.D., M.P. and I.S. processed clinical samples. C.O.B. performed negative-stain electron-microscopy analysis. T.Y.O., A.P.W. and V.R. performed bioinformatics analysis. D.F.R., P.D.B., P.J.B., T. Hatziioannou, C.M.R. and M.C.N. wrote the manuscript with input from all co-authors.

Competing interests The Rockefeller University has filed a provisional patent application in connection with this work on which D.F.R. and M.C.N. are inventors (US patent 63/021,387).

Additional information

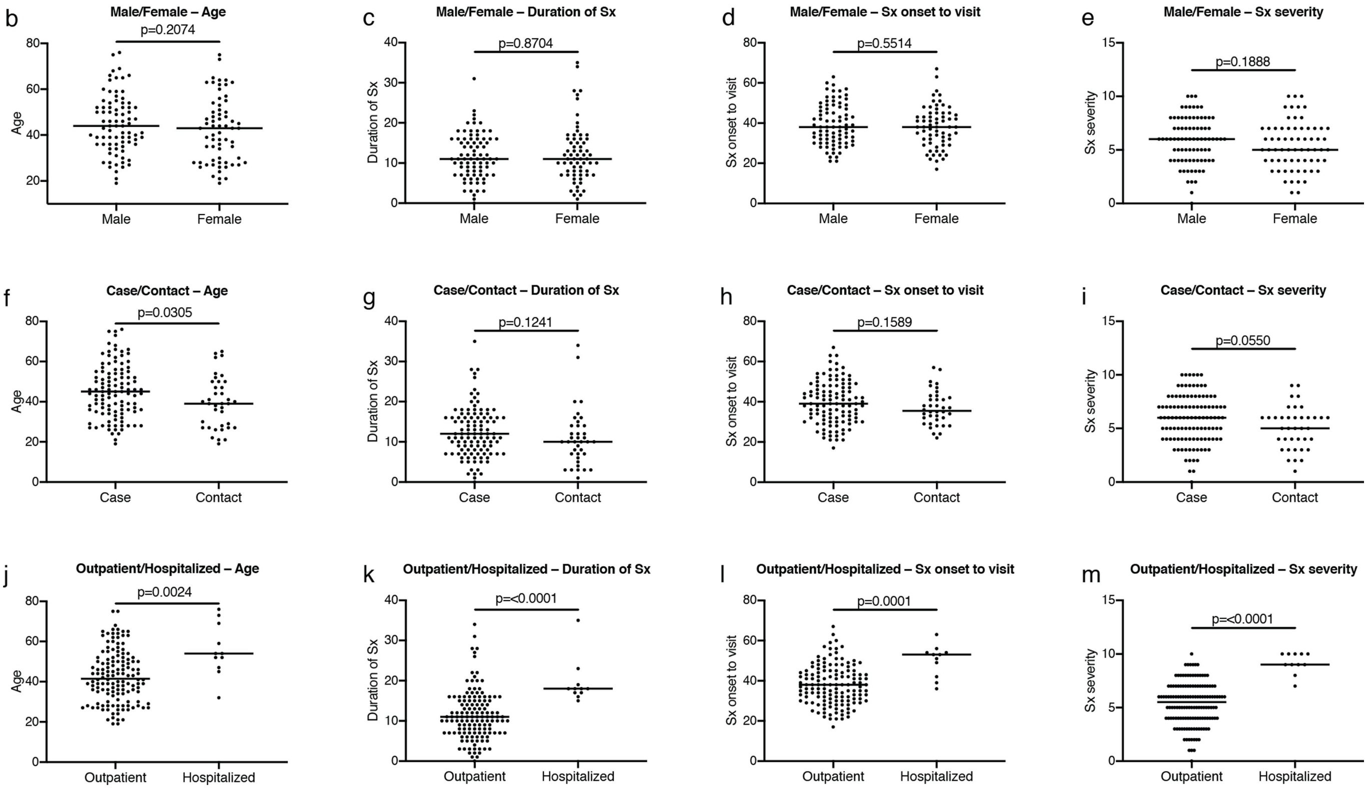
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2456-9.

Correspondence and requests for materials should be addressed to D.F.R., P.J.B., P.D.B., M.C. or M.C.N.

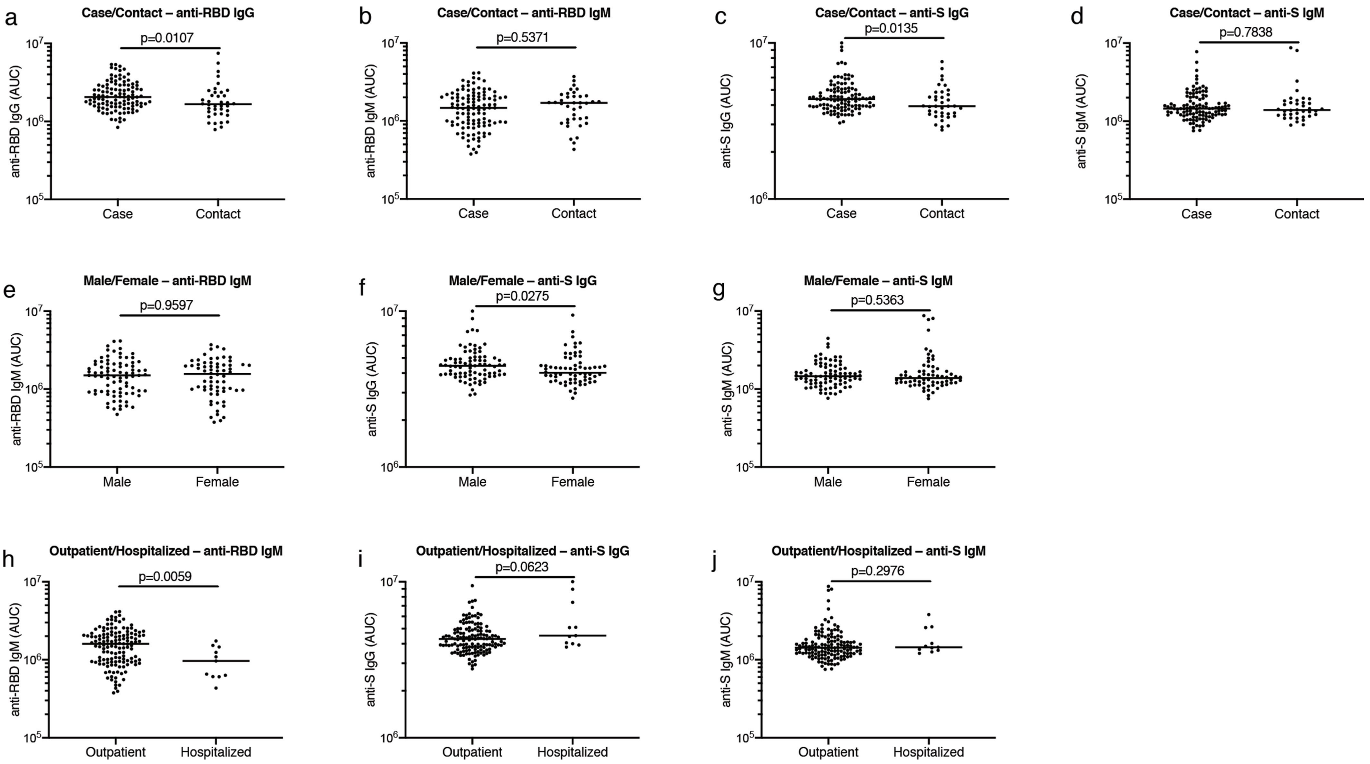
Peer review information *Nature* thanks Alice McHardy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

a

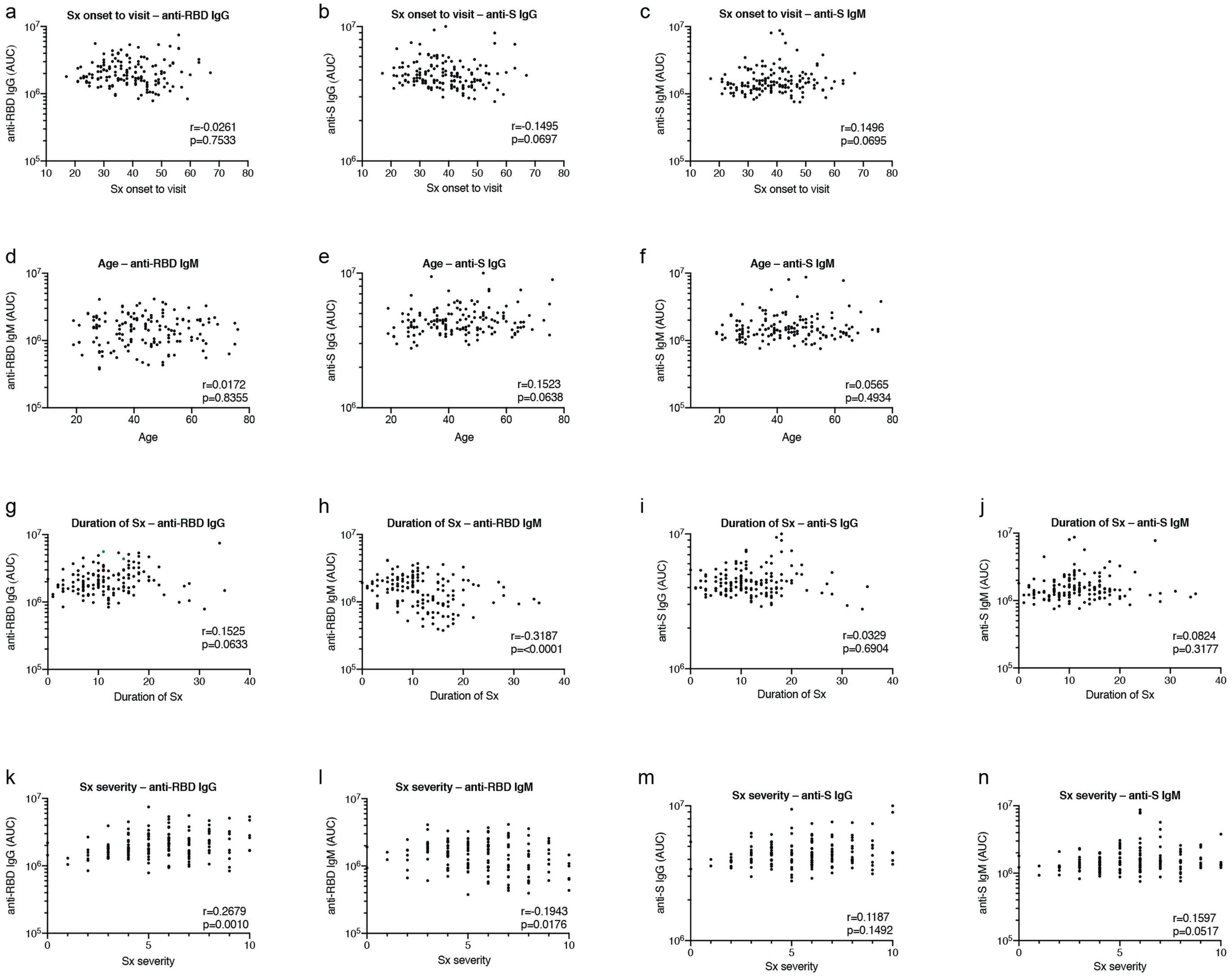
				Average duration		Average	ELISA binding (AUC)				
							RBD		S		
Gender	n	Average age	Case/Contact	Sx total	Sx onset to visit	Sx Severity (0-10)	IgG	IgM	IgG	IgM	Neutralization (NT50)
Male	83	45 (19-76)	65/18	12 (0-31)	39 (21-63)	5.8 (0-10)	2.44	1.61	4.65	1.62	867
Female	66	42 (19-75)	46/20	12 (1-35)	38 (17-67)	5.4 (1-9)	1.99	1.58	4.36	1.86	522



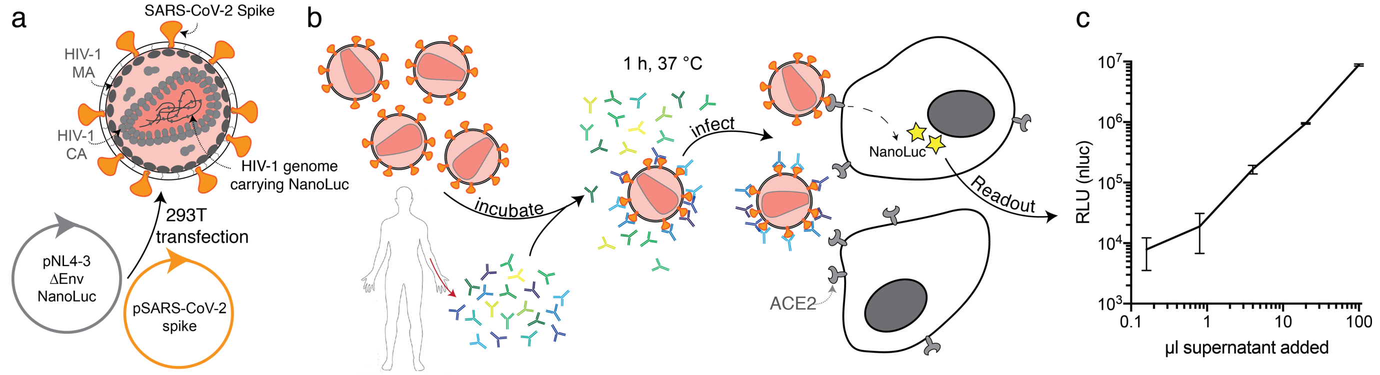
Extended Data Fig. 1 | Clinical correlates. a, Summary of the characteristics of the cohort. Sx, symptoms. **b**, Age distribution for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.2074. **c**, Duration of symptoms in days for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.8704. **d**, Time between symptom onset and plasma collection for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.5514. **e**, Subjective symptom severity on a scale of 0–10 for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.1888. **f**, Age distribution for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.0305. **g**, Duration of symptoms in days for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.1241. **h**, Time between symptom onset and plasma collection in days for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.1589. **i**, Symptom severity for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.0550. **j**, Age distribution for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants. *P* = 0.0024. **k**, Duration of symptoms in days for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* < 0.0001. **l**, Time between symptom onset and plasma collection in days for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* = 0.0001. **m**, Symptom severity for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* < 0.0001. Horizontal bars indicate median values. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests.



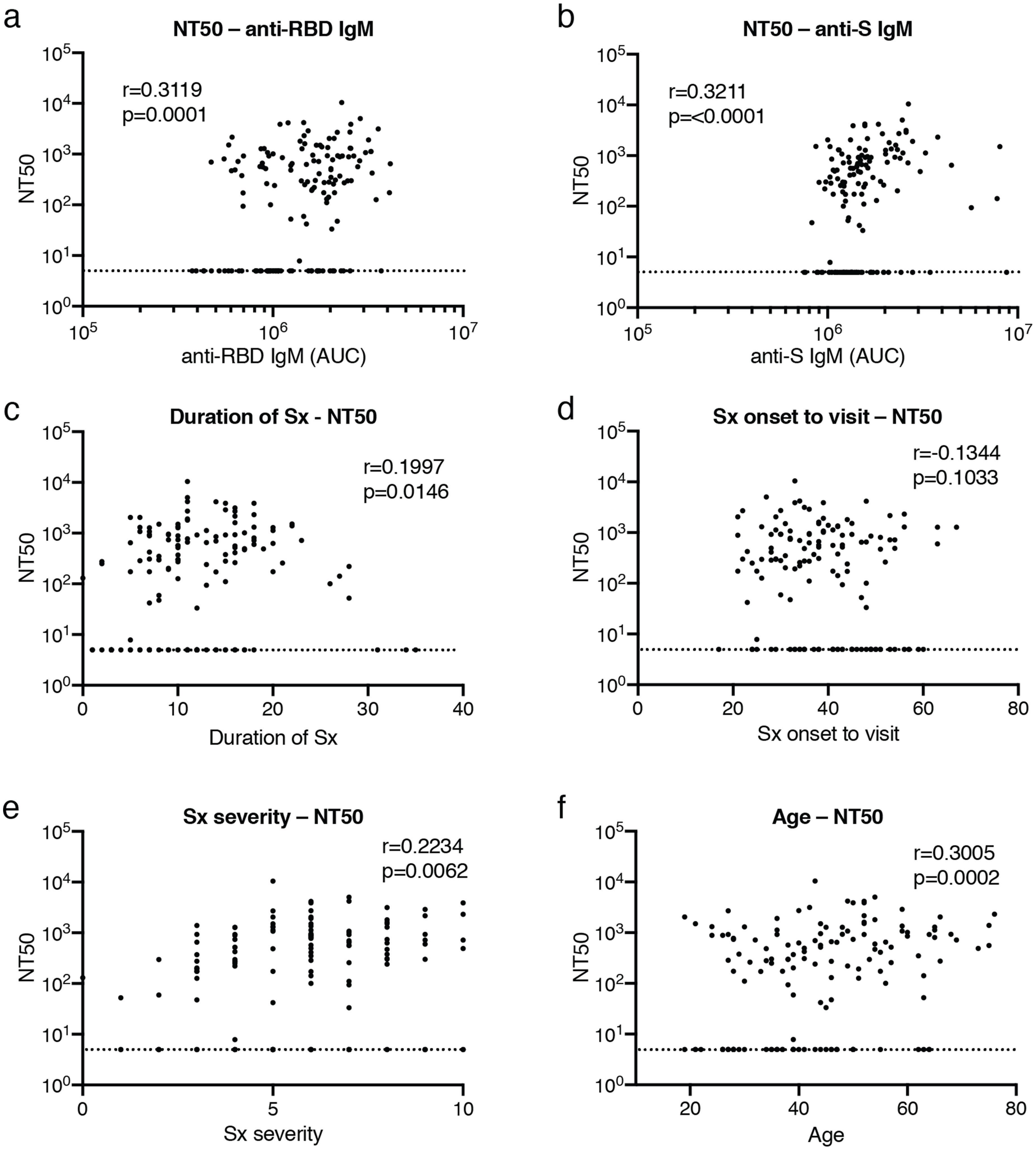
Extended Data Fig. 2 | Clinical correlates of plasma antibody titres. **a**, Normalized AUC for IgG anti-RBD for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.0107. **b**, Normalized AUC for IgM anti-RBD for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.5371. **c**, Normalized AUC for IgG anti-S for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.0135. **d**, Normalized AUC for IgM anti-S for all cases (*n* = 111) and contacts (*n* = 38) in the cohort. *P* = 0.7838. **e**, Normalized AUC for IgM anti-RBD for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.9597. **f**, Normalized AUC for IgG anti-S for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.0275. **g**, Normalized AUC for IgM anti-S for all men (*n* = 83) and women (*n* = 66) in the cohort. *P* = 0.5363. **h**, Normalized AUC for IgM anti-RBD for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* = 0.0059. **i**, Normalized AUC for IgG anti-S for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* = 0.0623. **j**, Normalized AUC for IgM anti-S for all outpatient (*n* = 138) and hospitalized (*n* = 11) participants in the cohort. *P* = 0.2976. Horizontal bars indicate median values. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests.



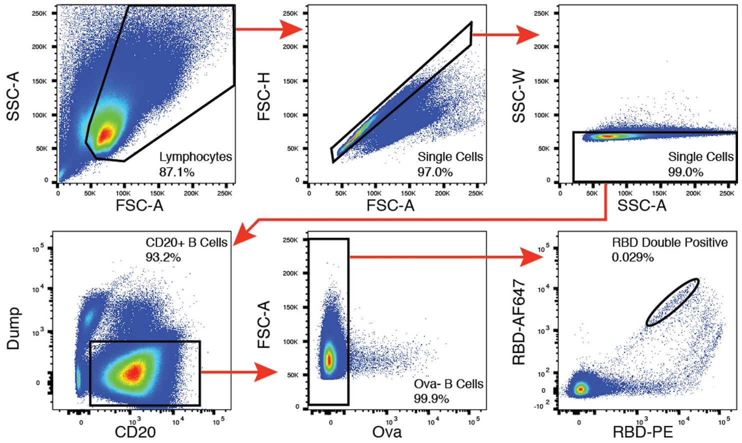
Extended Data Fig. 3 | Additional clinical correlates of plasma antibody titres. **a**, Time between symptom onset and plasma collection in days plotted against normalized AUC for IgG anti-RBD. *r* = −0.0261, *P* = 0.7533. **b**, Time between symptom onset and plasma collection in days plotted against normalized AUC for IgG anti-S. *r* = −0.1495, *P* = 0.0697. **c**, Time between symptom onset and plasma collection in days plotted against normalized AUC for IgM anti-S. *r* = 0.1496, *P* = 0.0695. **d**, Age plotted against AUC for IgM anti-RBD. *r* = 0.0172, *P* = 0.8355. **e**, Age plotted against normalized AUC for IgG anti-S. *r* = 0.1523, *P* = 0.0638. **f**, Age plotted against normalized AUC for IgM anti-S. *r* = 0.0565, *P* = 0.4934. **g**, Duration of symptoms in days plotted against normalized AUC for IgG anti-RBD. *r* = 0.1525, *P* = 0.0633. **h**, Duration of symptoms in days plotted against normalized AUC for IgM anti-RBD. *r* = −0.3187, *P* < 0.0001. **i**, Duration of symptoms in days plotted against normalized AUC for IgG anti-S. *r* = 0.0329, *P* = 0.6904. **j**, Duration of symptoms in days plotted against normalized AUC for IgM anti-S. *r* = 0.0824, *P* = 0.3177. **k**, Severity of symptoms plotted against normalized AUC for IgG anti-RBD. *r* = 0.2679, *P* = 0.0010. **l**, Severity of symptoms plotted against normalized AUC for IgM anti-RBD. *r* = −0.1943, *P* = 0.0176. **m**, Severity of symptoms plotted against normalized AUC for IgG anti-S. *r* = 0.1187 *P* = 0.1492. **n**, Severity of symptoms plotted against normalized AUC for IgM anti-S. *r* = 0.1597, *P* = 0.0517. All correlations were analysed by two-tailed Spearman's tests; *n* = 149.



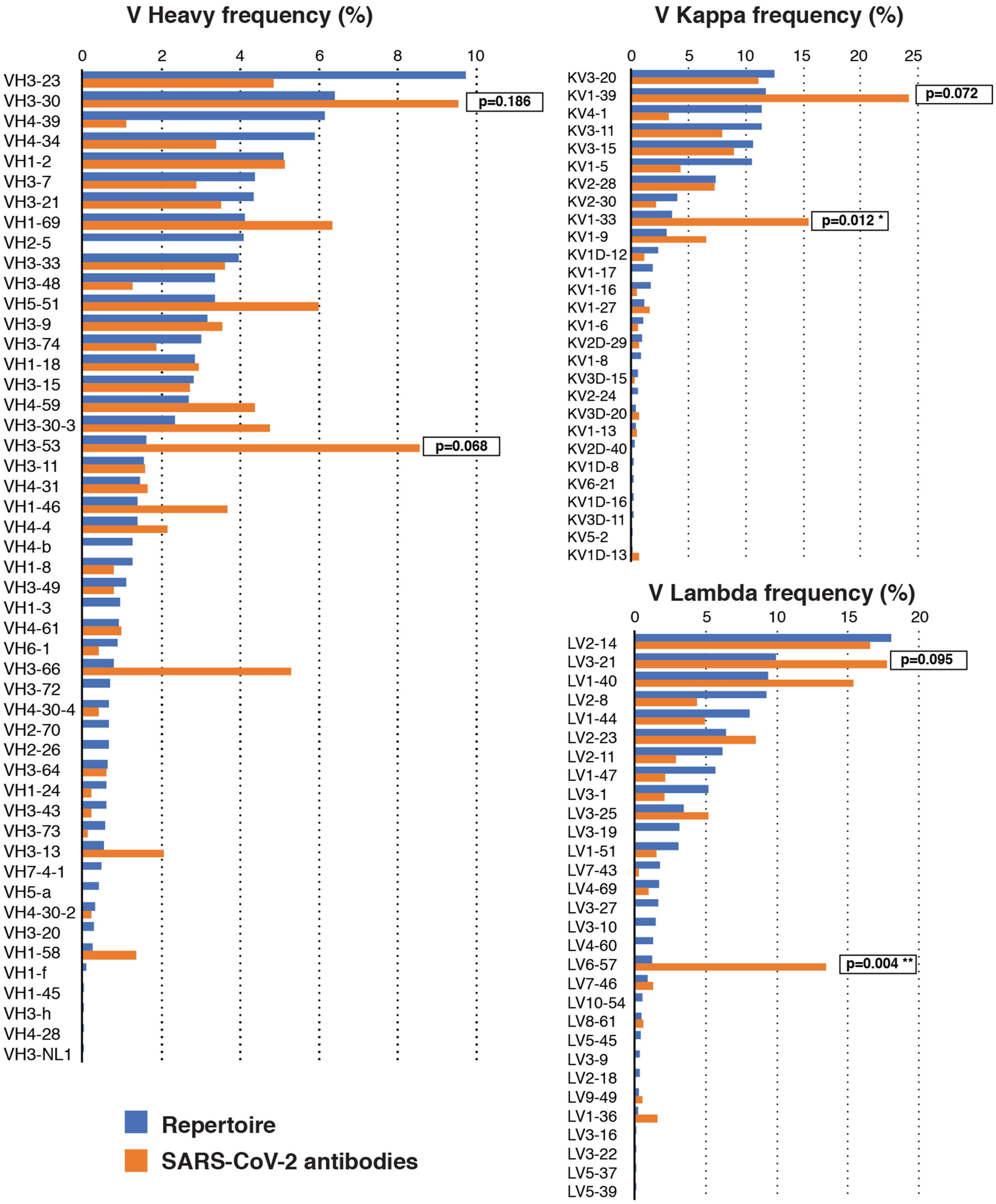
Extended Data Fig. 4 | Diagram of the SARS-CoV-2 pseudovirus luciferase assay. a, Co-transfection of pNL4-3ΔEnv-nanoluc and pSARS-CoV-2 spike vectors into 293T cells (ATCC) leads to production of SARS-CoV-2 spike-pseudotyped HIV-1 particles (SARS-CoV-2 pseudovirus) carrying the *nanoluc* gene. **b**, SARS-CoV-2 pseudovirus is incubated for 1 h at 37 °C with plasma or monoclonal antibody dilutions. The virus–antibody mixture is used to infect ACE2-expressing 293T cells, which will express nanoluc luciferase upon infection. **c**, Relative luminescence unit (RLU) reads from lysates of ACE2-expressing 293T cells infected with increasing amounts of SARS-CoV-2 pseudovirus. Data are mean ± s.d. of triplicates. One representative experiment is shown.



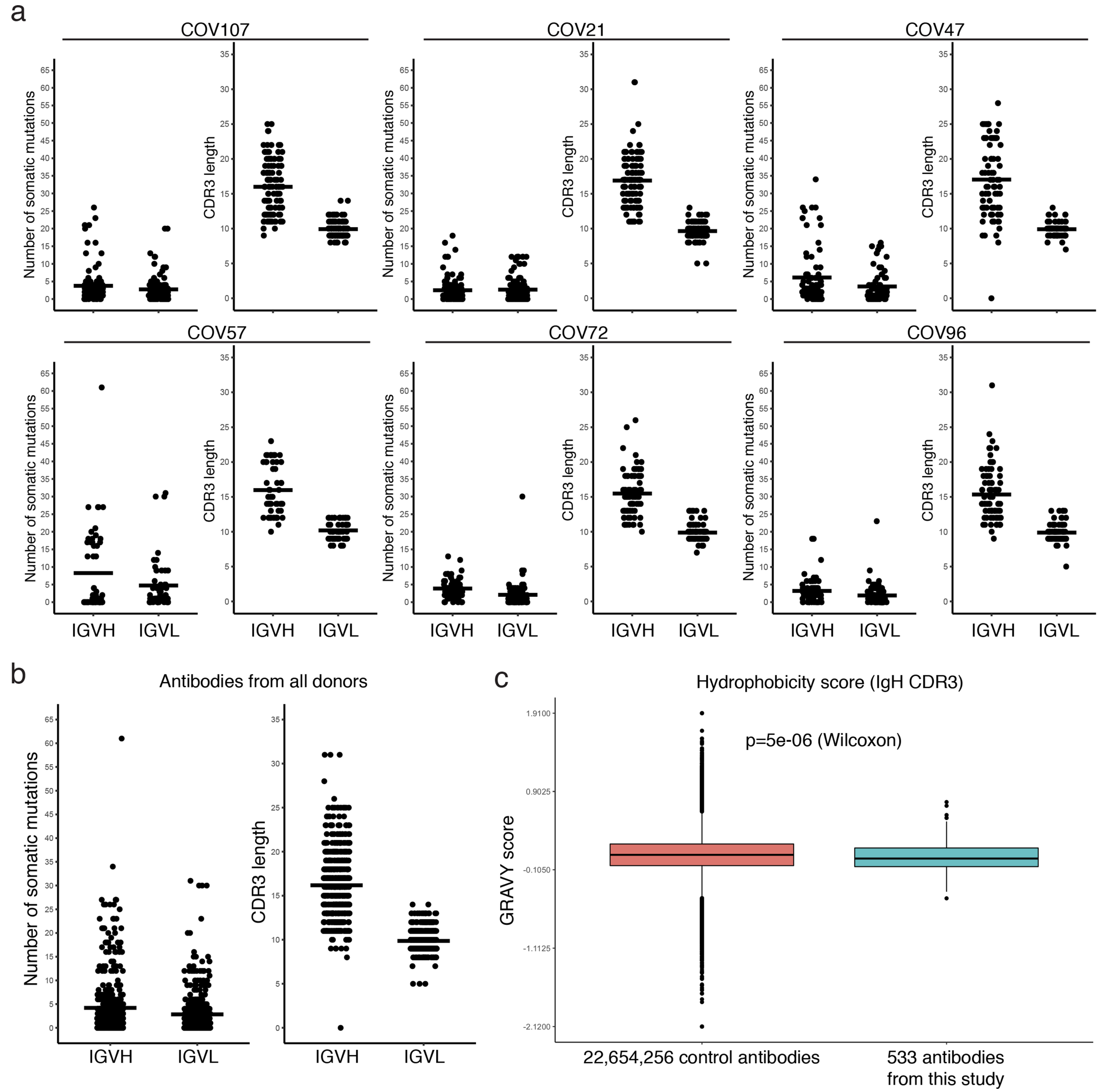
Extended Data Fig. 5 | Clinical correlates of neutralization. **a**, Normalized AUC for anti-RBD IgM plotted against NT_{50}. $r = 0.3119$, $P = 0.0001$. **b**, Normalized AUC for anti-S IgM plotted against NT_{50}. $r = 0.3211$, $P < 0.0001$. **c**, Duration of symptoms in days plotted against NT_{50}. $r = 0.1997$, $P = 0.0146$. **d**, Time between symptom onset and plasma collection in days plotted against NT_{50}. $r = -0.1344$, $P = 0.1033$. **e**, Symptom severity plotted against NT_{50}. $r = 0.2234$, $P = 0.0062$. **f**, Age plotted against NT_{50}. $r = 0.3005$, $P = 0.0002$. All correlations were analysed by two-tailed Spearman's tests; $n = 149$. The dotted line ($NT_{50} = 5$) represents the lower limit of detection of the pseudovirus neutralization assay. Samples with neutralizing titres below 1:50 were plotted at the lower limit of detection.



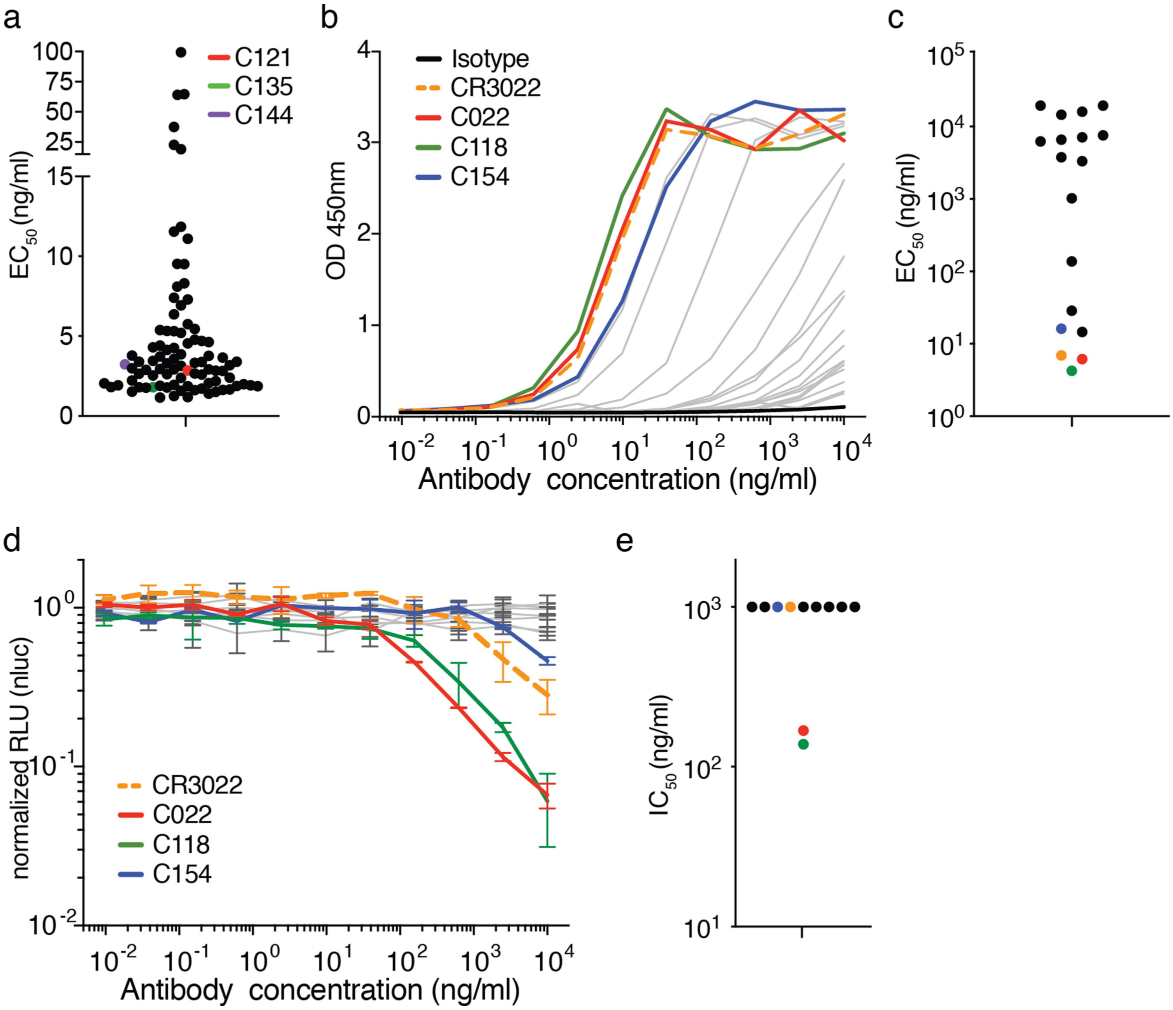
Extended Data Fig. 6 | Flow cytometry. Gating strategy used for cell sorting. Gating was on singlets that were CD20$^+$ and CD3$^-$CD8$^-$CD16$^-$Ova$^-$. Sorted cells were RBD-PE$^+$ and RBD-AF647$^+$.



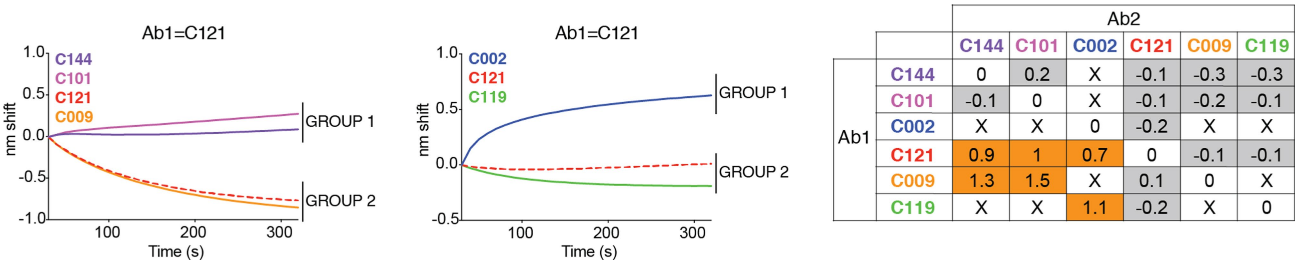
Extended Data Fig. 7 | Frequency distributions of human V genes. Two-tailed *t*-tests with unequal variance were used to compare the frequency distributions of human V genes of anti-SARS-CoV-2 antibodies from this study to Sequence Read Archive accession SRP010970[41].



Extended Data Fig. 8 | Analysis of antibody somatic hypermutation and CDR3 length. a, For each individual, the number of somatic nucleotide mutations at the IGVH and IGVL is shown on the left, and the amino acid length of the CDR3s is shown on the right. The horizontal bars indicate the mean. The number of antibody sequences (IGVH and IGVL) evaluated for each participant are *n* = 118 (COV107), *n* = 127 (COV21), *n* = 79 (COV47), *n* = 54 (COV57), *n* = 78 (COV72), *n* = 78 (COV96). **b**, Same as in **a** but for all antibodies combined (*n* = 534 for both IGVH and IGVL). **c**, Distribution of the hydrophobicity GRAVY scores at the IGH CDR3 in antibody sequences from this study compared to a public database (see Methods for statistical analysis). The box limits are at the lower and upper quartiles, the centre line indicates the median, the whiskers are 1.5× interquartile range and the dots represent outliers.



Extended Data Fig. 9 | Binding of the monoclonal antibodies to the RBD of SARS-CoV-2 and cross-reactivity to SARS-CoV. **a**, EC_{50} values for binding to the RBD of SARS-CoV-2. Average of two or more experiments. *n* = 89. **b**, **c**, Binding curves (**b**; representative experiment) and EC_{50} values (**c**; mean of two experiments) for binding to the RBD of SARS-CoV. *n* = 20 and *n* = 17 (excluding isotype and CR3022) for **b** and **c**, respectively. **d**, **e**, SARS-CoV pseudovirus neutralization curves and IC_{50} values. **d**, Data are mean ± s.d. of duplicates for one representative experiment. **e**, Data are the mean of two experiments (*n* = 10, excluding CR3022). Samples with IC_{50} values above 1 μg ml^{-1} were plotted at 1 μg ml^{-1}.



	Ab2					
Ab1	C144	C101	C002	C121	C009	C119
C144	0	0.2	X	-0.1	-0.3	-0.3
C101	-0.1	0	X	-0.1	-0.2	-0.1
C002	X	X	0	-0.2	X	X
C121	0.9	1	0.7	0	-0.1	-0.1
C009	1.3	1.5	X	0.1	0	X
C119	X	X	1.1	-0.2	X	0

Extended Data Fig. 10 | Biolayer interferometry experiment. Binding of antibodies C144, C101, C002, C121, C009 and C119. Graphs show second antibody binding to preformed C121 IgG–RBD complexes. The table displays the shift in nanometres after second antibody (Ab2) binding to the antigen in the presence of the first antibody (Ab1). Values are normalized through the subtraction of the autologous antibody control.

nature research

Corresponding author(s): Davide Robbiani, Paul Bieniasz, Marina Caskey, Pamela Bjorkman, Michel Nussenzweig

Last updated by author(s): Jun 9, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	IRIS by iMedRIS version 11.01 for clinical data collection and management; BD FACSDiva Software Version 8.0.2 for flow sorting; Omega version 5.11 by BMG Labtech for luminometer; Modulus II Microplate Reader User interface version 2.1.0 by TURNER BioSystems; MetaXpress V 6.1.2071 by Molecular Devices; SerialEM automated image acquisition software version 3.7.
Data analysis	FlowJo 10.6.2 for FACS analysis; GraphPad Prism 8.4.2; Microsoft Excel 16.36; MacVector 17.5.4 for sequence analysis; Omega MARS V2.10 by BMG Labtech for luminometer; Fortebio Octet Data Analysis Software 8.0; cryoSPARC v2.15 and UCSF chimera version 1.13.1 for EM analysis; code for sequence analysis can be obtained at https://github.com/stratust/igpipeline.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data are provided in SI Tables 1, 3, 4, 5 and 6; and Figure 3 has associated raw sequencing data (https://github.com/stratust/igpipeline).
Databases used in this study include A Public Database of Memory and Naive B-Cell Receptor Sequences (https://datadryad.org/stash/dataset/doi:10.5061/dryad.35ks2), PDB datasets 6VYB and 6NB6 and Sequence Read Archive SRP010970.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size of 157 individuals was based on how many we were able to recruit for blood donation between April 1 and May 8, 2020.
Data exclusions	8 contacts (i.e. exposed to SARS-CoV-2 confirmed infected individuals, but themselves not tested by RT-PCR) that did not develop symptoms were excluded from further analyses as it is possible that they were not infected. Exclusion criteria were not pre-established.
Replication	All experiments successfully repeated at least twice.
Randomization	This is not relevant as this is an observational study.
Blinding	This is not relevant as this is an observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Mouse anti-human CD20-PECy7 (BD Biosciences, 335793), clone L27 Mouse anti-human CD3-APC-eFluro 780 (Invitrogen, 47-0037-41), clone OKT3 Mouse anti-human CD8-APC-421eFluro 780 (Invitrogen, 47-0086-42), clone OKT8 Mouse anti-human CD16-APC-eFluro 780 (Invitrogen, 47-0168-41), clone eBioCB16 Mouse anti-human CD14-APC-eFluro 780 (Invitrogen, 47-0149-4), clone 61D3 Peroxidase Goat Anti-Human IgG Jackson Immuno Research 109-036-088 Peroxidase Goat Anti-Human IgM Jackson Immuno Research 109-035-129 Rabbit polyclonal anti-SARS-CoV-2 nucleocapsid antibody (catalog no. GTX135357; GeneTex) Goat anti-rabbit AlexaFluor 594 (catalog no. A-11012; Life Technologies) Anti-Zika virus monoclonal antibody Z021 (Robbiani et al, Cell 2017) used as isotype control
Validation	The human monoclonal antibody Z021, which binds to the Envelope Domain III of the Zika virus, was previously reported and validated (PMID: 31413072). No validation statements for the other antibodies that are commercially available.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293T (ATCC CRL-11268) 293TAce2 (derived from 293T); new cell line generated in this study VeroE6 (ATCC CRL-1586) Expi293F (ThermoFisher cat. A14527)

	Huh 7.5 (a derivative of Huh 7) was generated in the Laboratory of Virology and Infectious Disease, Rockefeller University (Dr. Charles Rice)
Authentication	Not authenticated after purchase, with the exception of the Huh 7.5 cells (authenticated by Genetica Cell Line Testing)
Mycoplasma contamination	The cells were checked for mycoplasma contamination by Hoechst staining or MycoAlert Kit from Lonza.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	We enrolled 83 males and 66 females with an average age of 45 and 42, respectively. Eligible participants were adults aged 18-76 years who were either diagnosed with SARS-CoV-2 infection by RT-PCR and were free of symptoms of COVID-19 for at least 14 days (cases), or who were close contacts (e.g., household, co-workers, members of same religious community) with someone who had been diagnosed with SARS-CoV-2 infection by RT-PCR and were free of symptoms suggestive of COVID-19 for at least 14 days (contacts). Exclusion criteria included presence of symptoms suggestive of active SARS-CoV-2 infection, or hemoglobin < 12 g/dL for males and < 11 g/dL for females.
Recruitment	Study participants were recruited at the Rockefeller University Hospital in New York from April 1 through May 8, 2020. Most study participants were residents of the Greater New York City tri-state region and were enrolled sequentially according to eligibility criteria. Participants were first interviewed by phone to collect information on their clinical presentation, and subsequently presented to the Rockefeller University Hospital for a single blood sample collection. The requirement for participants to be free of symptoms for at least 14 days might have favoured enrollment of participants that developed mild COVID-19 courses of infection during the first weeks of recruitment.
Ethics oversight	The Rockefeller University Institutional Review Board (1230 York Avenue, box 330, New York, NY 10065). Protocol DRO-1006 approved on February 6, 2020.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Whole blood samples were obtained from study participants recruited through Rockefeller University Hospital. Peripheral blood mononuclear cells (PBMCs) were separated by Ficoll gradient centrifugation. Prior to sorting, PBMCs were enriched for B cells using a Miltenyi Biotech pan B cell isolation kit (cat. no. 130-101-638) and LS columns (cat. no. 130-042-401).
Instrument	FACS Aria III (Becton Dickinson)
Software	BD FACSDiva Software Version 8.0.2 and FlowJo 10.6.2
Cell population abundance	Sorting efficiency ranged from 40% to 66%. This is calculated based on the number of IgG-specific antibody sequences that could be PCR-amplified successfully from single sorted cells from each donor.
Gating strategy	Cells were first gated for lymphocytes in FSC-A (x-axis) versus SSC-A (y-axis). We identify single cells in FSC-A versus FSC-H, and then SSC-A versus SSC-W. We then select for CD20+ Dump- B Cells in dump (anti-CD3-eFluro 780, anti-CD16-eFluro 780, anti-CD8-eFluro 780, anti-CD14-eFluro 780, Zombie NIR) versus CD20 (anti-CD20-PE-Cy7); dump-negative was considered to be signal less than 250, and CD20-positive was taken to be signal greater than 100. We then gate for Ova- B cells in FSC-A versus Ova-BV711; Ova-negative was considered to be all cells with signal less than 102. Select for TBEV double-positive cells in TBEV EDIII PE versus TBEV EDIII AlexaFluor 647; this gate was made along the 45° diagonal, above 103 on both axes. See also Extended Data Figure 6.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Potently neutralizing and protective human antibodies against SARS-CoV-2

https://doi.org/10.1038/s41586-020-2548-6

Received: 19 May 2019

Accepted: 7 July 2020

Seth J. Zost[1,15], Pavlo Gilchuk[1,15], James Brett Case[2], Elad Binshtein[1], Rita E. Chen[2,3], Joseph P. Nkolola[4], Alexandra Schäfer[5], Joseph X. Reidy[1], Andrew Trivette[1], Rachel S. Nargi[1], Rachel E. Sutton[1], Naveenchandra Suryadevara[1], David R. Martinez[5], Lauren E. Williamson[6], Elaine C. Chen[6], Taylor Jones[1], Samuel Day[1], Luke Myers[1], Ahmed O. Hassan[2], Natasha M. Kafai[2,3], Emma S. Winkler[2,3], Julie M. Fox[2], Swathi Shrihari[2], Benjamin K. Mueller[7], Jens Meiler[7,8], Abishek Chandrashekar[4], Noe B. Mercado[4], James J. Steinhardt[9], Kuishu Ren[10], Yueh-Ming Loo[10], Nicole L. Kallewaard[10], Broc T. McCune[2], Shamus P. Keeler[2,11], Michael J. Holtzman[2,11], Dan H. Barouch[4], Lisa E. Gralinski[5], Ralph S. Baric[5], Larissa B. Thackray[2], Michael S. Diamond[2,3,12,13], Robert H. Carnahan[1,14] ✉ & James E. Crowe Jr[1,6,14] ✉

The ongoing pandemic of coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a major threat to global health[1] and the medical countermeasures available so far are limited[2,3]. Moreover, we currently lack a thorough understanding of the mechanisms of humoral immunity to SARS-CoV-2[4]. Here we analyse a large panel of human monoclonal antibodies that target the spike (S) glycoprotein[5], and identify several that exhibit potent neutralizing activity and fully block the receptor-binding domain of the S protein (S_{RBD}) from interacting with human angiotensin-converting enzyme 2 (ACE2). Using competition-binding, structural and functional studies, we show that the monoclonal antibodies can be clustered into classes that recognize distinct epitopes on the S_{RBD}, as well as distinct conformational states of the S trimer. Two potently neutralizing monoclonal antibodies, COV2-2196 and COV2-2130, which recognize non-overlapping sites, bound simultaneously to the S protein and neutralized wild-type SARS-CoV-2 virus in a synergistic manner. In two mouse models of SARS-CoV-2 infection, passive transfer of COV2-2196, COV2-2130 or a combination of both of these antibodies protected mice from weight loss and reduced the viral burden and levels of inflammation in the lungs. In addition, passive transfer of either of two of the most potent ACE2-blocking monoclonal antibodies (COV2-2196 or COV2-2381) as monotherapy protected rhesus macaques from SARS-CoV-2 infection. These results identify protective epitopes on the S_{RBD} and provide a structure-based framework for rational vaccine design and the selection of robust immunotherapeutic agents.

The S protein of SARS-CoV-2 is the molecular determinant of viral attachment, fusion and entry into host cells[6]. The S protein is composed of an N-terminal subunit (S1) that mediates receptor binding, and a C-terminal subunit (S2) that mediates fusion between the virus and the membrane of the host cell. The S1 subunit contains an N-terminal domain (NTD) and a receptor-binding domain (RBD). SARS-CoV-2 and SARS-CoV, the genomes of which share approximately 78% sequence identity[1], both use human ACE2 as an entry receptor[7–9]. Human antibodies to the S glycoprotein mediate protective immunity against other zoonotic betacoronaviruses of high pathogenicity, including SARS-CoV[10–14] and Middle East respiratory syndrome coronavirus (MERS-CoV)[15–24]. The most potent S-protein-specific monoclonal antibodies appear to neutralize betacoronaviruses by binding to the region on the S_{RBD} that directly mediates receptor engagement, and thereby

---

[1]Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, TN, USA. [2]Department of Medicine, Washington University School of Medicine, St Louis, MO, USA. [3]Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA. [4]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [5]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [6]Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. [7]Department of Chemistry, Vanderbilt University, Nashville, TN, USA. [8]Leipzig University Medical School, Institute for Drug Discovery, Leipzig, Germany. [9]Antibody Discovery and Protein Engineering, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, USA. [10]Microbial Sciences, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, USA. [11]Division of Pulmonary and Critical Care Medicine, Washington University School of Medicine, St Louis, MO, USA. [12]Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO, USA. [13]Andrew M. and Jane M. Bursky Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, St Louis, MO, USA. [14]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA. [15]These authors contributed equally: Seth J. Zost, Pavlo Gilchuk.
✉e-mail: Robert.carnahan@vumc.org; james.crowe@vumc.org

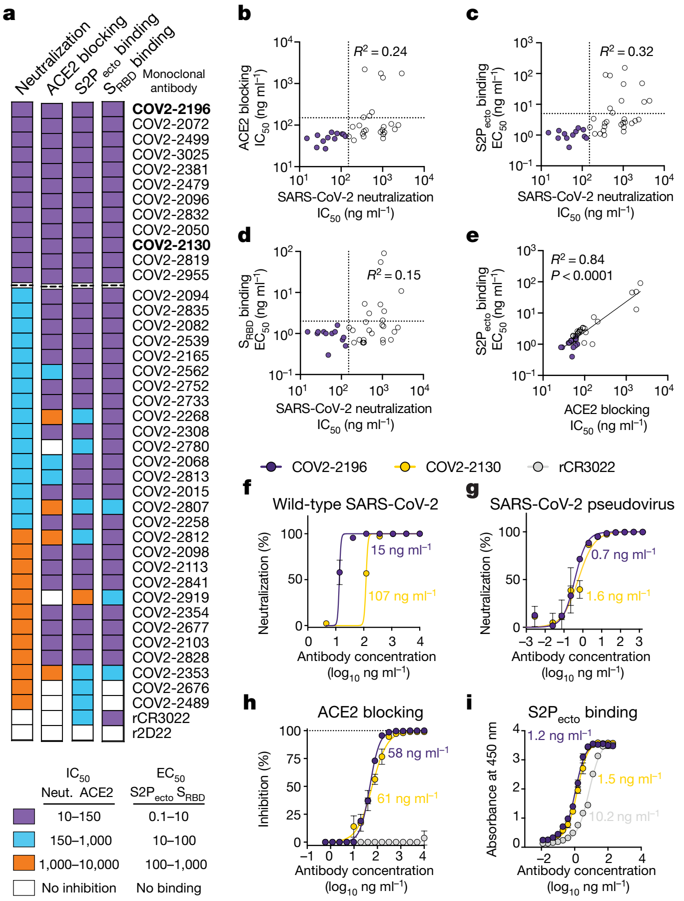


Fig. 1 | Functional characteristics of neutralizing SARS-CoV-2 monoclonal antibodies. a, Heat map of monoclonal antibody neutralization activity, human ACE2-blocking activity, and binding to either trimeric $S2P_{ecto}$ protein or monomeric S_{RBD}. Monoclonal antibodies are ordered by neutralization potency, and dashed lines indicate the 12 antibodies with a neutralization IC_{50} value <150 ng ml^{-1}. IC_{50} values (ng ml^{-1}) are shown for viral neutralization (neut.) and human ACE2 blocking, and EC_{50} values (ng ml^{-1}) for binding. The cross-reactive SARS-CoV S_{RBD} monoclonal antibody rCR3022 is shown as a positive control and the anti-dengue monoclonal antibody r2D22 as a negative control. Data are representative of at least two independent experiments performed in technical duplicate. No inhibition or no binding indicates an IC_{50} or EC_{50} value >10,000 ng ml^{-1}, respectively. **b**–**d**, Correlation of human ACE2 blocking (**b**), $S2P_{ecto}$ trimer binding (**c**) or S_{RBD} binding (**d**) of monoclonal antibodies with their neutralization activity. **e**, Correlation of human ACE2 blocking and $S2P_{ecto}$ trimer binding. R^2 values are shown for linear regression analysis of log-transformed values. Purple circles indicate monoclonal antibodies with a neutralization IC_{50} value <150 ng ml^{-1}. **f**, Neutralization curves for COV2-2196 and COV2-2130 against wild-type SARS-CoV-2 virus. Calculated IC_{50} values are shown on the graph. Error bars, s.d.; data are representative of at least two independent experiments performed in technical duplicate. **g**, Neutralization curves for COV2-2196 and COV2-2130 in a pseudovirus neutralization assay. Error bars, s.d.; values are technical duplicates from a single experiment. Calculated IC_{50} values from a minimum of six experiments are shown on the graph. **h**, Human-ACE2-blocking curves for COV2-2196, COV2-2130 and rCR3022 in a human-ACE2-blocking ELISA. Calculated IC_{50} values are shown on the graph. Data are mean ± s.d. of technical triplicates from a representative experiment repeated twice. **i**, ELISA binding of COV2-2196, COV2-2130 and rCR3022 to trimeric $S2P_{ecto}$. Calculated EC_{50} values are shown on the graph. Data are mean ± s.d. of technical triplicates from a representative experiment repeated twice.

blocking the attachment of the virus to host cells. Human antibodies could be used for prophylaxis, post-exposure prophylaxis or treatment of SARS-CoV-2 infection[25]. Many studies are ongoing—including randomized controlled trials assessing plasma from convalescent individuals with prior SARS-CoV-2 infection, and one trial evaluating hyperimmune immunoglobulin—but it is not yet clear whether these treatments can reduce morbidity or mortality[26].

We isolated 389 SARS-CoV-2 S-protein-reactive monoclonal antibodies from the B cells of two convalescing individuals who had been infected with SARS-CoV-2 in Wuhan, China[5]. A subset of those antibodies bound to a recombinant RBD construct (S_{RBD}) and exhibited neutralizing activity in a rapid screening assay with wild-type SARS-CoV-2 virus[5]. In the current study, we sought to define the antigenic landscape of SARS-CoV-2 and determine which sites of the S_{RBD} are targets of neutralizing monoclonal antibodies. We tested 40 of the anti-S human monoclonal antibodies that were previously pre-selected by rapid neutralization screening assay in a quantitative focus reduction neutralization test (FRNT) with the WA1/2020 strain of SARS-CoV-2. The antibodies in our panel of 40 exhibited half-maximum inhibitory concentration (IC_{50}) values that ranged from 15 to over 4,000 ng ml^{-1} (visualized as a heat map in Fig. 1a, values shown in Supplementary Table 1 and full curves shown in Extended Data Fig. 1). We hypothesized that many of these S_{RBD}-reactive monoclonal antibodies neutralize virus infection by blocking the binding of the S_{RBD} to human ACE2. Indeed, most of the neutralizing monoclonal antibodies that we tested inhibited the interaction of human ACE2 with trimeric S protein directly (Fig. 1a, Extended Data Fig. 2). Consistent with these results, these monoclonal antibodies also bound strongly to a trimeric S ectodomain ($S2P_{ecto}$) protein or to monomeric S_{RBD} (Fig. 1a, Extended Data Fig. 3). We evaluated whether the potency of the antibodies at binding $S2P_{ecto}$ or S_{RBD} or blocking human ACE2 predicted binding neutralization potency independently, but none of these measurements correlated with neutralization potency (Fig. 1b–d). However, the antibodies within the highest neutralizing potency tier of the panel (IC_{50} < 150 ng ml^{-1}) also had the strongest blocking activity against human ACE2 (IC_{50} < 150 ng ml^{-1}) and exceptional binding activity (half-maximum effective concentration (EC_{50}) < 2 ng ml^{-1}) to the $S2P_{ecto}$ trimer (Fig. 1e). Representative neutralization curves for two potently neutralizing monoclonal antibodies designated COV2-2196 and COV2-2130 are shown in Fig. 1f. Potent neutralization was confirmed using pseudovirus neutralization assays, which revealed far-more sensitive neutralization phenotypes than the wild-type virus and demonstrated a requirement for the use of live virus in assays for assessment of monoclonal antibody potency (Fig. 1g). Both of these monoclonal antibodies (COV2-2196 and COV2-2130) bound strongly to the $S2P_{ecto}$ trimer and fully blocked the binding of human ACE2 (Fig. 1h, i).

We next defined the major antigenic sites on the S_{RBD} for neutralizing monoclonal antibodies by competition-binding analysis. We first used a biolayer-interferometry-based competition assay with a minimal version of the S_{RBD} domain to screen for monoclonal antibodies that competed for binding with the potently neutralizing monoclonal antibody COV2-2196 or a recombinant version of the previously described SARS-CoV monoclonal antibody CR3022, which recognizes a conserved cryptic epitope[12,27]. We identified three major groups of competing monoclonal antibodies (Fig. 2a). The largest group of antibodies blocked COV2-2196 but not recombinant CR3022 (rCR3022), whereas some monoclonal antibodies were blocked by rCR3022 but not by COV2-2196. Two monoclonal antibodies, including COV2-2130, were not blocked by either reference monoclonal antibody. Most monoclonal antibodies competed with human ACE2 for binding, suggesting that they bound near the ACE2-binding site of the S_{RBD}. We used COV2-2196, COV2-2130 and rCR3022 in an enzyme-linked immunosorbent assay (ELISA)-based competition-binding assay with the $S2P_{ecto}$ trimer and found that the S_{RBD} contained three major antigenic sites, with some monoclonal antibodies probably making contacts in more than one site (Fig. 2b). Most of the potently neutralizing monoclonal antibodies directly competed with COV2-2196 for binding. Competition-binding analyses of human ACE2 and monoclonal antibodies with serum or

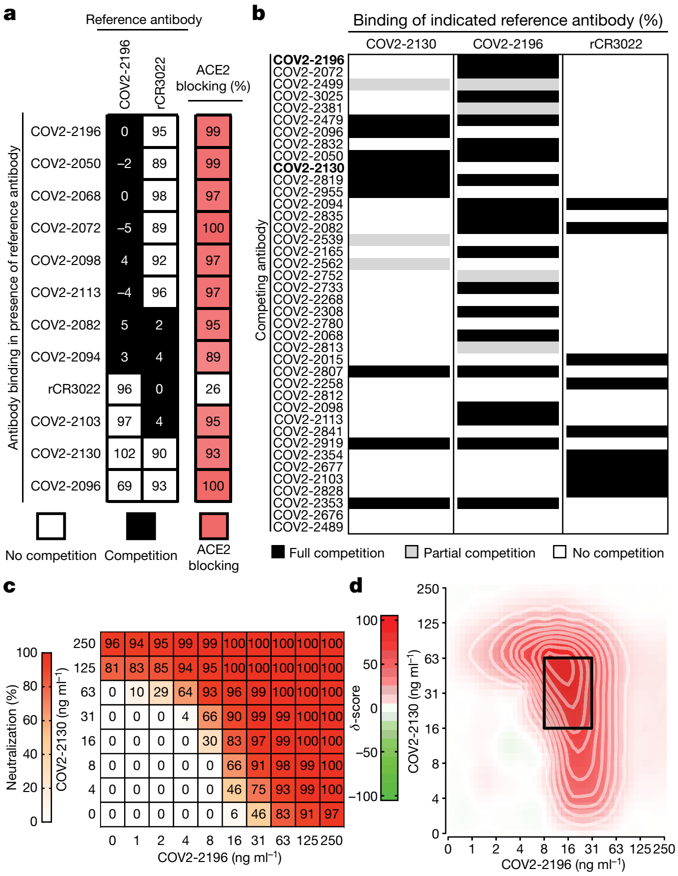


Fig. 2 | Epitope mapping of monoclonal antibodies by competition-binding analysis and synergistic neutralization by a pair of monoclonal antibodies. **a**, Left, monoclonal antibody binding to the S_{RBD} in the presence of reference monoclonal antibodies COV2-2196 or rCR3022. Values in squares are the per cent binding of the monoclonal antibody in the presence of the competing monoclonal antibody relative to a mock-competition control. Black squares, full competition (<33% relative binding); white squares, no competition (>67% relative binding). Right, biolayer-interferometry-based competition binding assay measuring the ability of monoclonal antibodies to prevent the binding of human ACE2. Values are the per cent blocking of human ACE2 by the monoclonal antibody. Red indicates high blocking activity. **b**, Competition of the panel of neutralizing monoclonal antibodies with reference monoclonal antibodies COV2-2130, COV2-2196 or rCR3022. Binding of reference monoclonal antibodies to trimeric $S2P_{ecto}$ was measured in the presence of saturating competitor monoclonal antibody in a competition ELISA and normalized to binding in the presence of r2D22. Black, full competition (<25% binding of reference antibody); grey, partial competition (25–60% binding of reference antibody); white, no competition (>60% binding of reference antibody). **c**, Neutralization dose–response matrix of wild-type SARS-CoV-2 by COV2-2196 and COV2-2130. Axes denote the concentration of each monoclonal antibody, with the per cent neutralization shown in each square. Data are from a representative experiment performed in technical triplicate and repeated twice. The white-to-red heat map denotes 0% neutralization to 100% neutralization, respectively. **d**, Synergy map calculated on the basis of the SARS-CoV-2 neutralization in **c**. δ-score is a synergy score. Red colour indicates areas in which synergistic neutralization was observed; black box indicates the area of maximum synergy between the two monoclonal antibodies.

plasma from four previously described individuals with recent laboratory-confirmed SARS-CoV-2 infection[5] showed that COV2-2196- and COV2-2130-like antibody responses are subdominant in these individuals (Extended Data Fig. 4).

As COV2-2196 and COV2-2130 did not compete for binding to the S_{RBD}, we assessed whether these monoclonal antibodies synergize for virus neutralization—a phenomenon that has been observed previously for SARS-CoV monoclonal antibodies[12]. We tested combination responses (Fig. 2c) in an FRNT using SARS-CoV-2, and compared the values obtained experimentally with the expected responses calculated by synergy-scoring models[28]. The comparison revealed that the combination of COV2-2196 and COV2-2130 antibodies was synergistic, with an overall synergy δ-score of 17.4 (where any score greater than 10 indicates synergy; Fig. 2d). In particular, a combined monoclonal antibody dose of 79 ng ml^{-1} (16 ng ml^{-1} of COV2-2196 and 63 ng ml^{-1} of COV2-2130) had the same activity as 250 ng ml^{-1} of each individual antibody (Fig. 2c). This finding shows that by using a cocktail of two antibodies, the dose of each antibody can be reduced by more than threefold to achieve the same potency of virus neutralization in vitro.

We next defined the epitopes that are recognized by representative monoclonal antibodies in the two major competition-binding groups that synergize for neutralization. We used mutagenesis to determine critical residues in the S_{RBD} for the binding of neutralizing monoclonal antibodies (Fig. 3a, Extended Data Fig. 5). These studies showed that F486 or N487 are critical residues for the binding of COV2-2196, and N487 is a critical residue for COV2-2165—two antibodies that compete with one another for binding. Likewise, mutagenesis studies for COV2-2130 using K444A and G447R mutants suggested that these residues (K444 and G447) are critical for recognition (Fig. 3a). Previous structural studies have defined the interaction between the S_{RBD} and human ACE2[29] (Fig. 3b). Most of the interacting residues in the S_{RBD} are contained within a 60-amino-acid linear peptide that defines the human ACE2 recognition motif (Fig. 3c). We next tested the binding of human monoclonal antibodies to this minimal peptide and found that potent neutralizing members of the largest group of antibodies from the competition-binding assay—including COV2-2196, COV2-2165 and COV2-2832—recognized this peptide (Fig. 3c), suggesting that these monoclonal antibodies make critical contacts within the human ACE2 recognition motif.

We used negative-stain electron microscopy of the $S2P_{ecto}$ trimer in complex with antigen-binding fragments (Fabs) to determine the structural epitopes for several monoclonal antibodies (Fig. 3d, e, Supplementary Table 2). The potently neutralizing antibodies COV2-2196 and COV2-2165 bound to the human ACE2 recognition motif of the S_{RBD} and recognized the 'open' conformational state of the $S2P_{ecto}$ trimer, in which the RBD rotates upward to expose the residues that mediate ACE2 interaction[30,31] (Fig. 3d). COV2-2130, which represents a different competition-binding group, bound to the RBD in the $S2P_{ecto}$ trimer in the 'closed' position (Fig. 3d). Because COV2-2196 and COV2-2130 did not compete for binding, we attempted to make complexes of both Fabs bound at the same time to the $S2P_{ecto}$ trimer. We found that both Fabs bound simultaneously when the $S2P_{ecto}$ trimer was in the open position, indicating that COV2-2130 can recognize the S_{RBD} in both conformations (Fig. 3e). Overlaying the structure of the two-Fab complex with that of the S_{RBD}–CR3022 complex[27], we observed that these antibodies bind to three distinct sites on the S_{RBD}, as predicted by our competition-binding studies (Fig. 3f).

Next, we tested the prophylactic efficacy of COV2-2196 or COV2-2130 monotherapy or a combination of both COV2-2196 and COV2-2130 in a model of SARS-CoV-2 infection in BALB/c mice. In this model (Fig. 4a), mice are first treated with an anti-IFNAR1 antibody and then transduced with an adenovirus that expresses human ACE2 (AdV-hACE2), which results in susceptibility to infection with SARS-CoV-2, viral replication and severe bronchopneumonia[32]. The mice were treated with a single dose of COV2-2196 or COV2-2130, a cocktail of COV2-2196 and COV2-2130, or an isotype control monoclonal antibody one day before intranasal challenge with a 4×10^5 plaque-forming unit (PFU) dose of SARS-CoV-2. Prophylaxis with COV2-2196, COV2-2130 or their combination prevented severe SARS-CoV-2-induced weight loss in the mice during the first week of infection (Fig. 4b). Viral RNA levels were reduced significantly at 7 days post-infection (dpi) in the lung and in distant sites including the heart and spleen (Fig. 4c). The expression of cytokine and

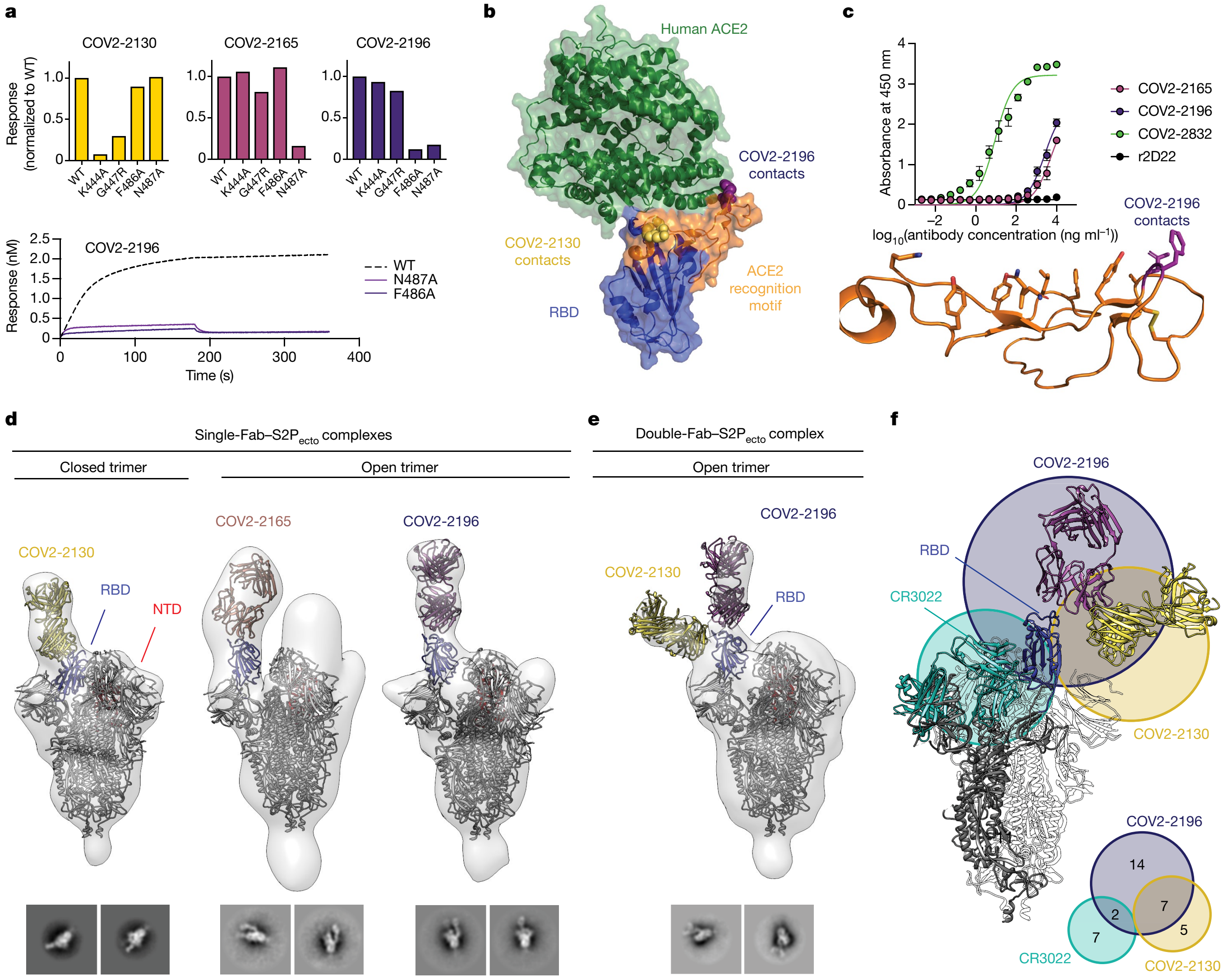


Fig. 3 | Epitope identification and structural characterization of monoclonal antibodies. a, Identification of critical contact residues by alanine and arginine mutagenesis. Top, binding of COV2-2130, COV2-2165 or COV2-2196 to wild-type (WT) or mutant S_{RBD} constructs, normalized to the wild type. Bottom, representative binding curves for COV2-2196 to wild-type or mutant S_{RBD} constructs. **b**, Co-crystal structure of the S_{RBD} (blue) and human ACE2 (green) (Protein Data Bank (PDB): 6M0J), with the human ACE2 recognition motif shown in orange. Critical contact residues are shown for COV2-2130 (gold spheres) and COV2-2196 (purple spheres). **c**, Top, ELISA binding of monoclonal antibodies to the human ACE2 recognition motif. r2D22 is shown as a negative control. Data are mean ± s.d. of technical triplicates from a single experiment repeated twice. Bottom, structure of human ACE2 recognition motif (orange) with COV2-2196 critical contact residues shown (purple). **d**, Fab–$S2P_{ecto}$ trimer complexes visualized by negative-stain electron microscopy for COV2-2130, COV2-2165 and COV2-2196. The S_{RBD} is shown in blue, the S_{NTD} in red and electron density in grey. The trimer state (open or closed) is denoted for each complex. Representative two-dimensional (2D) class averages for each complex are shown at the bottom (box size is 128 pixels, with 3.06 Å per pixel). Data are from a single experiment; detailed collection statistics are provided in Supplementary Table 2. **e**, COV2-2130 and COV2-2196 Fabs in complex with the $S2P_{ecto}$ trimer. Colours and data collection as in **d**. Representative 2D class averages for the complexes are shown at the bottom; scales as in **d**. **f**, Top, the competition-binding landscape visualized on the $S2P_{ecto}$ trimer. The CR3022 crystal structure was docked into the double-Fab–$S2P_{ecto}$ trimer model. CR3022 is shown in cyan. Bottom, quantitative Venn diagram showing the number of monoclonal antibodies in each competition group.

chemokine genes—indicative of inflammation—was also reduced in the lungs of each group of COV2-antibody-treated mice at 7 dpi (Fig. 4d).

We also tested COV2-2196, COV2-2130 and their combination for prophylactic efficacy in an immunocompetent model using a mouse-adapted SARS-CoV-2 (MA-SARS-CoV-2) virus[33] (Fig. 4e, f). Each of the monoclonal antibody treatments reduced viral RNA levels by up to 10^5-fold at 2 dpi in the lung, compared to the isotype control group (Fig. 4f). All of the mice from the COV2-2196 and the combined COV2-2196 and COV2-2130 treatment groups, and 8 out of 10 mice from the COV2-2130 treatment group, no longer had infectious virus in the lung at 2 dpi (as measured by a plaque assay of lung tissue homogenates; Fig. 4f).

We evaluated the effect of treatment with monoclonal antibodies on SARS-CoV-2-induced lung pathology. At 7 dpi, lungs from anti-IFNAR1-treated, AdV-hACE2-transduced mice that were treated with isotype control monoclonal antibody and then inoculated with SARS-CoV-2- showed perivascular, peribronchial and alveolar inflammation, with the infiltration of immune cells and alveolar damage that are characteristic of viral pneumonia (Fig. 4g, Supplementary Table 3). By contrast, mice under the same conditions that were treated with COV2-2196, COV2-2130 or their combination developed notably less lung disease, and their lung pathology was similar to that observed in AdV-hACE2-transduced control mice that were not infected with SARS-CoV-2 (Fig. 4g, Supplementary Table 3).

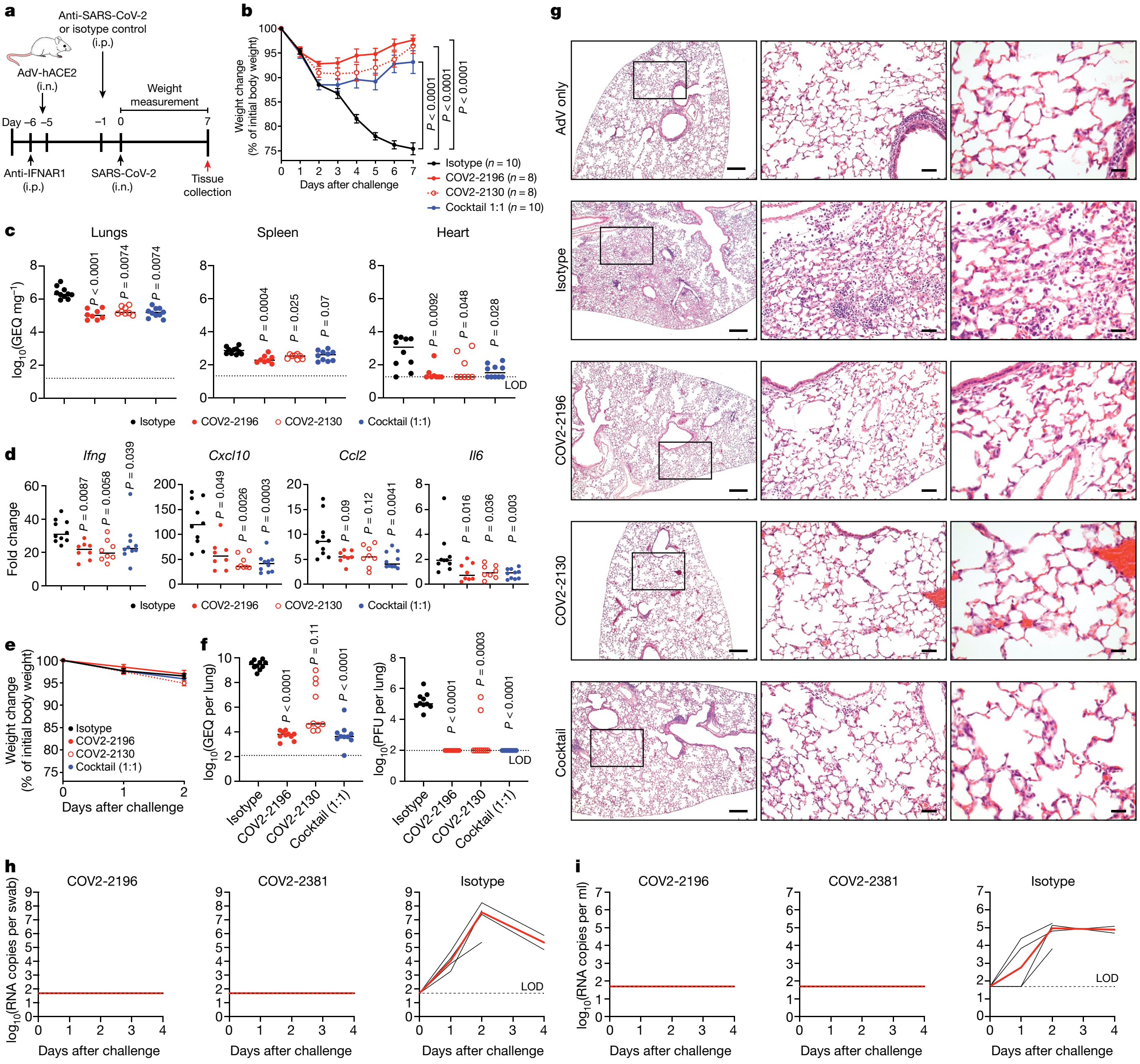


Fig. 4 | Prophylactic efficacy of neutralizing human monoclonal antibodies against SARS-CoV-2 infection in mouse and NHP models in vivo. **a**, SARS-CoV-2 challenge model. Mice were treated with anti-IFNAR1 and transduced with AdV-hACE2 followed by the passive transfer of 200 μg of COV2-2196, COV-2130, their combination (1:1 ratio) or an isotype control monoclonal antibody (i.n., intranasal; i.p., intraperitoneal). One day later, mice were inoculated intranasally with SARS-CoV-2. Tissues were collected at 7 dpi for analysis (**c**, **d**). **b**, Body weight change of mice in **a** with comparison to isotype control using a repeated measurements two-way analysis of variance (ANOVA) with Tukey's post hoc test. Data are mean ± s.e.m. of each experimental group. The number of mice (*n*) for each experimental group is shown. **c**, **d**, Viral burden (measured as $\log_{10}$(number of genome equivalents (GEQ) per mg)) at 7 dpi in the lungs, spleen and heart (**c**) and the expression of cytokine and chemokine genes (**d**) were measured by RT–qPCR assay. Comparisons were performed using a Kruskal–Wallis ANOVA with Dunn's post hoc test. **e**, **f**, MA-SARS-CoV-2 challenge model. Mice were treated with the indicated monoclonal antibody and then inoculated intranasally with MA-SARS-CoV-2. **e**, Body weight change of mice (mean ± s.e.m. of each experimental group; $n = 10$ mice per group). **f**, Viral burden at 2 dpi in the lungs, measured by RT–qPCR (left) or plaque assay (right) from **e**; comparisons were made using a Kruskal–Wallis ANOVA with Dunn's post hoc test ($n = 10$ mice per group). **g**, Haematoxylin and eosin staining of lung sections from mice that were treated and challenged as in **a**, shown at day 7. Images are shown at low (left), medium (middle) and high (right) magnification. Each image is representative of two separate experiments ($n = 3$ to 5 mice per group). Scale bars, 250 μm (left); 50 μm (middle); 25 μm (right). **h**, **i**, SARS-CoV-2 NHP challenge model. Rhesus macaques received one 50 mg kg^{-1} dose of COV2-2196 ($n = 4$ macaques per group), COV2-2381 ($n = 4$ macaques per group) or isotype control monoclonal antibody ($n = 4$ macaques per group) intravenously on day −3 and were then challenged intranasally and intratracheally with SARS-CoV-2 after three days. Subgenomic viral RNA levels were assessed in nasal swabs (**h**) and bronchioalveolar lavage (**i**) at multiple time points after challenge. Each black curve shows an individual macaque, with red lines indicating the median values within each treatment group. Data represent a single experiment. Dashed lines indicate the limit of detection (LOD) of the assay.

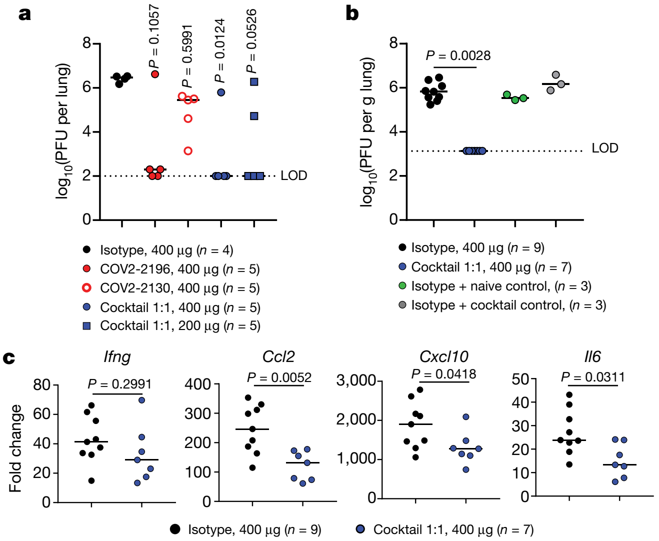


Fig. 5 | Therapeutic efficacy of neutralizing human monoclonal antibodies against SARS-CoV-2 infection. a, Mice were inoculated intranasally with MA-SARS-CoV-2 and 12 hours later given the indicated monoclonal antibody treatments by intraperitoneal injection. Viral burden in the lungs at 2 dpi was measured by plaque assay. The number of mice per group (*n*) is indicated. Data represent one experiment. **b**, Mice were treated with anti-IFNAR1 and transduced with AdV-hACE2. Mice were then inoculated intranasally with wild-type SARS-CoV-2 and 12 hours later given the indicated monoclonal antibody treatments by intraperitoneal injection. Viral burden in the lungs at 2 dpi was measured by plaque assay. Two experiments were performed with *n* = 3 to 5 mice per group. Controls for plaque neutralization assay performance were included: lung homogenates from individual mice (*n* = 3) that were treated with isotype control monoclonal antibody were mixed 1:1 (v:v) with lung homogenates from individual naive untreated mice or antibody-cocktail-treated mice. The latter mixture ensures that neutralization of infection did not occur ex vivo after tissue homogenization. For **a**, **b**, measurements from individual mice and median titre are shown, and each group was compared to the isotype-control-treated group using a Kruskal–Wallis ANOVA with Dunn's post hoc test. **c**, Expression of cytokine and chemokine genes was measured by qPCR analysis in lungs from **b**. Measurements from individual mice and median values are shown. Groups were compared using the two-sided Mann–Whitney *U*-test. The number of mice per group (*n*) is indicated. Two experiments were performed with *n* = 3 to 5 mice per group.

We next tested the protective efficacy of monoclonal antibodies using a recently described non-human primate (NHP) model of SARS-CoV-2[34,35]. In this model, we tested two monoclonal antibodies as monotherapy: COV2-2196 and another of the most potent antibodies identified, COV2-2381—a neutralizing monoclonal antibody that is encoded by the same variable gene segments as COV2-2196 but which contains a number of amino acid differences in the heavy-chain complementarity-determining region 3 (HCDR3) and light-chain complementarity-determining region 3 (LCDR3) (Extended Data Fig. 6a). Notably, other groups have identified highly similar monoclonal antibodies from multiple donors, demonstrating that these monoclonal antibodies constitute a public clonotype[36]. Rhesus macaques received one 50 mg kg^{-1} dose of COV2-2196, COV2-2381 or isotype control monoclonal antibody intravenously on day −3, and were then challenged intranasally and intratracheally on day 0 with a 1.1×10^4 PFU dose of SARS-CoV-2. After the challenge, we used quantitative PCR with reverse transcription (RT–qPCR) to quantify the levels of subgenomic viral RNA generated by viral replication in the bronchoalveolar lavage and in nasal swabs. High levels of subgenomic viral RNA were observed in the macaques that were treated with isotype control monoclonal antibody, with a median peak of 7.53 (range 5.37–8.23) RNA copies per swab in nasal swabs and 4.97 (3.81–5.24) log$_{10}$ RNA copies per ml in the bronchoalveolar lavage (Fig. 4h, i). Subgenomic viral RNA was not detected in samples from either of the antibody-treated groups (limit of detection = 50 (1.7 log$_{10}$) RNA copies per swab or per ml), indicating that these antibodies conferred protection against SARS-CoV-2. A pharmacokinetics analysis showed that the concentrations of circulating human monoclonal antibodies were similar in macaques from each treatment group (Extended Data Fig. 6b).

We next assessed the therapeutic efficacy of treatment with COV2-2196, COV2-2130 or their combination using the MA-SARS-CoV-2 mouse model. All treatments reduced the levels of infectious virus in the lungs of mice at 2 dpi. The antibody cocktail (1:1) delivered at a dose of 400 µg per mouse (around 20 mg kg^{-1}) was the most efficient; this treatment significantly reduced the viral burden in the lung by up to 3×10^4-fold, and four out of five mice from this treatment group did not have detectable levels of infectious virus in the lung (Fig. 5a). Similarly, treatment of AdV-hCE2-transduced mice with 400 µg per mouse of the cocktail 12 hours after challenge with wild-type SARS-CoV-2 virus revealed that infectious virus was fully neutralized in the lungs in vivo (Fig. 5b). Inflammation was also reduced in the lungs of mice that were treated with the antibody cocktail compared to the lungs of isotype-control-treated mice (Fig. 5c). Collectively, these in vivo results suggest that either of the potently neutralizing monoclonal antibodies COV2-2196 or COV2-2381 alone, and the combination of both COV2-2196 and COV2-2130, are promising candidates for the prevention or treatment of COVID-19.

Since the start of the SARS-CoV-2 pandemic, several groups have identified human monoclonal antibodies that bind to the S$_{\mathrm{RBD}}$ and neutralize the virus[36–44]. Here, we have defined the antigenic landscape for a number of potently neutralizing monoclonal antibodies against SARS-CoV-2 that were derived from a larger panel of hundreds of antibodies[5]. These studies demonstrate that although a wide range of human neutralizing antibodies are elicited by natural infection with SARS-CoV-2, only a small subset of those monoclonal antibodies are of high potency (IC$_{50}$ < 50 ng ml^{-1} against wild-type SARS-CoV-2 virus). Biochemical and structural analysis of these potent monoclonal antibodies defined three principal antigenic sites of vulnerability on the S$_{\mathrm{RBD}}$ for SARS-CoV-2 neutralization. Representative monoclonal antibodies from two antigenic sites were shown to synergize in vitro and confer protection as an in vivo cocktail in both prophylactic and therapeutic treatment. Our findings reveal critical features of effective humoral immunity to SARS-CoV-2 and suggest that the role of synergistic neutralization activity in polyclonal responses should be investigated further. Moreover, as SARS-CoV-2 continues to circulate, population-level immunity elicited by natural infection may start to select for antigenic variants that escape the selective pressure of neutralizing antibodies. Other groups have reported the selection of SARS-CoV-2 RBD escape mutations in the presence of single monoclonal antibodies, but not in the presence of a mixture of two antibodies[45], which reinforces the need to target multiple epitopes of the S protein in vaccines or immunotherapies. So far, the gene that encodes the S protein has been found to be limited in diversity—with the exception of a D614G substitution[46], which is far away from the amino acid positions identified in our mutational studies for the antibodies we have considered here. Rationally selected therapeutic cocktails such as the one we describe are likely to offer greater resistance to SARS-CoV-2 escape than single antibodies. Our results provide a basis for the preclinical evaluation and development of the identified monoclonal antibodies as candidates for use as COVID-19 immunotherapeutic agents in humans.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2548-6.

1. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med*. **382**, 727–733 (2020).
3. Tse, L. V., Meganck, R. M., Graham, R. L. & Baric, R. S. The current and future state of vaccines, antivirals and gene therapies against emerging coronaviruses. *Front. Microbiol*. **11**, 658 (2020).
4. Siracusano, G., Pastori, C. & Lopalco, L. Humoral immune responses in COVID-19 patients: a window on the state of the art. *Front. Immunol*. **11**, 1049 (2020).
5. Zost, S. J. et al. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat. Med*. https://doi.org/10.1038/s41591-020-0998-x (2020).
6. Pillay, T. S. Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *J. Clin. Pathol*. **73**, 366–369 (2020).
7. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol*. **94**, e00127-20 (2020).
8. Hoffmann, M., et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
9. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
10. Sui, J. et al. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc. Natl Acad. Sci. USA* **101**, 2536–2541 (2004).
11. ter Meulen, J. et al. Human monoclonal antibody as prophylaxis for SARS coronavirus infection in ferrets. *Lancet* **363**, 2139–2141 (2004).
12. ter Meulen, J. et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med*. **3**, e237 (2006).
13. Zhu, Z. et al. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **104**, 12123–12128 (2007).
14. Rockx, B. et al. Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. *J. Virol*. **82**, 3220–3235 (2008).
15. Chen, Z. et al. Human neutralizing monoclonal antibody inhibition of Middle East respiratory syndrome coronavirus replication in the common marmoset. *J. Infect. Dis*. **215**, 1807–1815 (2017).
16. Choi, J. H. et al. Characterization of a human monoclonal antibody generated from a B-cell specific for a prefusion-stabilized spike protein of Middle East respiratory syndrome coronavirus. *PLoS One* **15**, e0232757 (2020).
17. Niu, P. et al. Ultrapotent human neutralizing antibody repertoires against Middle East respiratory syndrome coronavirus from a recovered patient. *J. Infect. Dis*. **218**, 1249–1260 (2018).
18. Wang, L. et al. Importance of neutralizing monoclonal antibodies targeting multiple antigenic sites on the Middle East respiratory syndrome coronavirus spike glycoprotein to avoid neutralization escape. *J. Virol*. **92**, e02002-17 (2018).
19. Wang, N., et al. Structural definition of a neutralization-sensitive epitope on the MERS-CoV S1-NTD. *Cell Rep*. **28**, 3395–3405 (2019).
20. Zhang, S. et al. Structural definition of a unique neutralization epitope on the receptor-binding domain of MERS-CoV spike glycoprotein. *Cell Rep*. **24**, 441–452 (2018).
21. Corti, D. et al. Prophylactic and postexposure efficacy of a potent human monoclonal antibody against MERS coronavirus. *Proc. Natl Acad. Sci. USA* **112**, 10473–10478 (2015).
22. Jiang, L. et al. Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. *Sci. Transl. Med*. **6**, 234ra59 (2014).
23. Tang, X. C. et al. Identification of human neutralizing antibodies against MERS-CoV and their role in virus adaptive evolution. *Proc. Natl Acad. Sci. USA* **111**, E2018–E2026 (2014).
24. Ying, T. et al. Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies. *J. Virol*. **88**, 7796–7805 (2014).
25. Jiang, S., Hillyer, C. & Du, L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol*. **41**, 355–359 (2020).
26. Valk, S. J. et al. Convalescent plasma or hyperimmune immunoglobulin for people with COVID-19: a rapid review. *Cochrane Database Syst. Rev*. **5**, CD013600 (2020).
27. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
28. Ianevski, A., Giri, A. K. & Aittokallio, T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res*. **48**, W488–W493 (2020).
29. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
30. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
31. Walls, A.C., et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292 (2020).
32. Hassan, A. O. et al. A SARS-CoV-2 infection model in mice demonstrates protection by neutralizing antibodies. *Cell* https://doi.org/10.1016/j.cell.2020.06.011 (2020).
33. Dinnon, K. H. et al. A mouse-adapted SARS-CoV-2 model for the evaluation of COVID-19 medical countermeasures. Preprint at bioRxiv https://doi.org/10.1101/2020.05.06.081497 (2020).
34. Chandrashekar, A. et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* https://doi.org/10.1126/science.abc4776 (2020).
35. Yu, J. et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* https://doi.org/10.1126/science.abc6284 (2020).
36. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals. *Nature* https://doi.org/10.1038/s41586-020-2456-9 (2020).
37. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* https://doi.org/10.1126/science.abc5902 (2020).
38. Cao, Y. et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84 (2020).
39. Ju, B. et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* https://doi.org/10.1038/s41586-020-2380-z (2020).
40. Rogers, T. F. et al. Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *Science* https://doi.org/10.1126/science.abc7520 (2020).
41. Shi, R. et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).
42. Wec, A. Z. et al. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* https://doi.org/10.1126/science.abc7424 (2020).
43. Wu, Y. et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).
44. Hansen, J. et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* https://doi.org/10.1126/science.abd0827 (2020).
45. Baum, A. et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* https://doi.org/10.1126/science.abd0831 (2020).
46. Laha, S. et al. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol*. **85**, 104445 (2020).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and, with the exception of pathology scoring, the investigators were not blinded to allocation during experiments and outcome assessment.

Antibodies

The human antibodies studied in this paper were isolated from blood samples from two individuals in North America with previous laboratory-confirmed symptomatic SARS-CoV-2 infection that was acquired in China. The original clinical studies to obtain specimens after written informed consent were previously described[5] and had been approved by the Institutional Review Board of Vanderbilt University Medical Center, the Institutional Review Board of the University of Washington and the Research Ethics Board of the University of Toronto. The individuals (a 56-year-old male and a 56-year-old female) are a married couple and residents of Wuhan, China who travelled to Toronto, Canada, where PBMCs were obtained by leukopheresis 50 days after symptom onset. The antibodies were isolated using diverse tools for isolation and cloning of single antigen-specific B cells and the antibody variable genes that encode monoclonal antibodies[5].

Cell culture

Vero E6 (ATCC, CRL-1586), Vero (ATCC, CCL-81), HEK293 (ATCC, CRL-1573) and HEK293T (ATCC, CRL-3216) cells were maintained at 37 °C in 5% CO_2 in Dulbecco's minimal essential medium (DMEM) containing 10% (v/v) heat-inactivated fetal bovine serum (FBS), 10 mM HEPES pH 7.3, 1 mM sodium pyruvate, 1 × non-essential amino acids and 100 U ml^{-1} of penicillin–streptomycin. Vero-furin cells were obtained from T. Pierson and have been described previously[47]. FreeStyle 293F cells (Thermo Fisher Scientific, R79007) were maintained at 37 °C in 8% CO_2. Expi293F cells (Thermo Fisher Scientific, A1452) were maintained at 37 °C in 8% CO_2 in Expi293F Expression Medium (Thermo Fisher Scientific, A1435102). ExpiCHO cells (Thermo Fisher Scientific, A29127) were maintained at 37 °C in 8% CO_2 in ExpiCHO Expression Medium (Thermo Fisher Scientific, A2910002). Authentication analysis was not performed for the cell lines used. Mycoplasma testing of Expi293F and ExpiCHO cultures was performed on a monthly basis using a PCR-based mycoplasma detection kit (ATCC, 30-1012K).

Viruses

SARS-CoV-2 strain 2019 n-CoV/USA_WA1/2020 was obtained from the Centers for Disease Control and Prevention (a gift from N. Thornburg). Virus was passaged in Vero CCL81 cells and titrated by plaque assay on Vero E6 cells. MA-SARS-CoV-2 virus was generated as described previously[33]. Virus was propagated in Vero E6 cells grown in DMEM supplemented with 10% Fetal Clone II and 1% penicillin–streptomycin. The virus titre was determined by plaque assay. In brief, virus was diluted serially and inoculated onto confluent monolayers of Vero E6 cells, followed by an agarose overlay. Plaques were visualized on day 2 post-infection after staining with neutral red dye. All work with infectious SARS-CoV-2 was approved by the Washington University School of Medicine or UNC Chapel Hill Institutional Biosafety Committees and conducted in approved BSL3 facilities using appropriate powered air-purifying respirators and personal protective equipment.

Recombinant antigens and proteins

A gene encoding the ectodomain of a prefusion conformation-stabilized SARS-CoV-2 spike ($S2P_{ecto}$) protein was synthesized and cloned into a DNA plasmid expression vector for mammalian cells. A similarly designed S protein antigen with two prolines and removal of the furin cleavage site for stabilization of the prefusion form of S was reported previously[30]. In brief, this gene includes the ectodomain of SARS-CoV-2 (to residue 1,208), a T4 fibritin trimerization domain, an AviTag site-specific biotinylation sequence and a C-terminal 8×His tag. To stabilize the construct in the prefusion conformation, we included substitutions K986P and V987P and mutated the furin cleavage site at residues 682–685 from RRAR to ASVG. This recombinant spike 2P-stabilized protein (designated here as $S2P_{ecto}$) was isolated by metal affinity chromatography on HisTrap Excel columns (GE Healthcare), and protein preparations were purified further by size-exclusion chromatography on a Superose 6 Increase 10/300 column (GE Healthcare). The presence of trimeric, prefusion conformation S protein was verified by negative-stain electron microscopy[5]. For electron microscopy with S and Fabs, we expressed a variant of $S2P_{ecto}$ lacking an AviTag but containing a C-terminal Twin-Strep-tag, similar to that described previously[30]. Expressed protein was isolated by metal affinity chromatography on HisTrap Excel columns (GE Healthcare), followed by further purification on a StrepTrap HP column (GE Healthcare) and size-exclusion chromatography on TSKgel G4000SW_{XL} (TOSOH). To express the S_{RBD} subdomain of the SARS-CoV-2 S protein, residues 319–541 were cloned into a mammalian expression vector downstream of an IL-2 signal peptide and upstream of a thrombin cleavage site, an AviTag and a 6×His tag. RBD protein fused to the mouse IgG1 Fc domain (designated RBD–mFc), was purchased from Sino Biological (40592-V05H). For epitope mapping by alanine scanning, wild-type SARS-CoV-2 RBD (residues 334–526) or RBD single-mutation variants were cloned with an N-terminal CD33 leader sequence and C-terminal GSSG linker, AviTag, GSSG linker and 8×His tag. Spike proteins were expressed in FreeStyle 293 cells (Thermo Fisher Scientific) or Expi293 cells (Thermo Fisher Scientific) and isolated by affinity chromatography using a HisTrap column (GE Healthcare), followed by size-exclusion chromatography with a Superdex200 column (GE Healthcare). Purified proteins were analysed by SDS–PAGE to ensure purity and appropriate molecular weights.

Electron microscopy stain grid preparation, imaging and processing of SARS-CoV-2 $S2P_{ecto}$ protein or $S2P_{ecto}$–Fab complexes

To perform electron microscopy imaging, Fabs were produced by digesting recombinant chromatography-purified IgGs using resin-immobilized cysteine protease enzyme (FabALACTICA, Genovis). The digestion occurred in 100 mM sodium phosphate and 150 mM NaCl pH 7.2 (PBS) for around 16 h at ambient temperature. To remove cleaved Fc and intact IgG, the digestion mix was incubated with CaptureSelect Fc resin (Genovis) for 30 min at ambient temperature in PBS buffer. If needed, the Fab was buffer-exchanged into Tris buffer by centrifugation with a Zeba spin column (Thermo Fisher Scientific).

For screening and imaging of negatively stained SARS-CoV-2 $S2P_{ecto}$ protein in complex with human Fabs, the proteins were incubated at a molar ratio of 4 Fab:3 spike monomer for around 1 hour and approximately 3 µl of the sample at concentrations of about 10–15 µg ml^{-1} was applied to a glow-discharged grid with continuous carbon film on 400 square mesh copper electron microscopy grids (Electron Microscopy Sciences). The grids were stained with 0.75% uranyl formate[48]. Images were recorded on a Gatan US4000 4k × 4k CCD camera using an FEI TF20 (TFS) transmission electron microscope operated at 200 keV and control with SerialEM[49]. All images were taken at 50,000× magnification with a pixel size of 2.18 Å per pixel in low-dose mode at a defocus of 1.5–1.8 µm. The total dose for the micrographs was around 25–38 e^- per $Å^2$. Image processing was performed using the cryoSPARC software package[50]. Images were imported, and particles were CTF-estimated. The images were then denoised and picked with Topaz[51,52]. The particles were extracted with a box size of 256 pixels and binned to 128 pixels. 2D class averages were performed and good classes selected for ab initio model and refinement without symmetry. For electron microscopy model docking of SARS-CoV-2 $S2P_{ecto}$ protein, the closed model (PDB: 6VXX) was used in Chimera[53] for docking to the electron microscopy map (see also Supplementary Table 2 for details). For the SARS-CoV-2

S2P$_{ecto}$–Fab COV2-2165 and SARS-CoV-2 S2P$_{ecto}$–Fab COV2-2196 complexes, the open model of SARS-CoV-2 (PDB: 6VYB) and Fab (PDB: 12E8) was used in Chimera for docking to the electron microscopy maps (see also Supplementary Table 2 for details). For the SARS-CoV-2 S2P$_{ecto}$–Fab COV2-2130 complex, the closed model and Fab (PDB: 12E8) were used in Chimera for docking to the electron microscopy map (see also Supplementary Table 2 for details). All images were made with Chimera. PyMOL (Schrödinger) was used to visualize previously solved molecular structures of the SARS-CoV-2 RBD–human ACE2 complex and the 60-amino-acid human ACE2 recognition motif (PDB: 6M0J).

Monoclonal antibody production and purification

Sequences of monoclonal antibodies that had been synthesized (Twist Bioscience) and cloned into an IgG1 monocistronic expression vector (designated as pTwist-mCis_G1) were used for monoclonal antibody secretion in mammalian cell culture. This vector contains an enhanced 2A sequence and GSG linker that allows the simultaneous expression of monoclonal antibody heavy and light chain genes from a single construct upon transfection[54]. We previously described microscale expression of monoclonal antibodies in 1 ml ExpiCHO cultures in 96-well plates[5]. For larger-scale monoclonal antibody expression, we performed transfection (1–300 ml per antibody) of CHO cell cultures using the Gibco ExpiCHO Expression System and protocol for 50 ml mini bioreactor tubes (Corning) as described by the vendor. Culture supernatants were purified using HiTrap MabSelect SuRe (Cytiva, formerly GE Healthcare Life Sciences) on a 24-column parallel protein chromatography system (Protein BioSolutions). Purified monoclonal antibodies were buffer-exchanged into PBS, concentrated using Amicon Ultra-4 50-kDa centrifugal filter units (Millipore Sigma) and stored at 4 °C until use. Purified monoclonal antibodies were tested routinely for endotoxin levels (found to be less than 30 EU per mg IgG for mouse studies and less than 1 EU per mg IgG for NHP studies). Endotoxin testing was performed using the PTS201F cartridge (Charles River), with a sensitivity range from 10 to 0.1 EU per ml, and an Endosafe Nexgen-MCS instrument (Charles River).

ELISA binding assays

Wells of 96-well microtitre plates were coated with purified recombinant SARS-CoV-2 S protein or SARS-CoV-2 S$_{RBD}$ protein at 4 °C overnight. Plates were blocked with 2% non-fat dry milk and 2% normal goat serum in Dulbecco's phosphate-buffered saline (DPBS) containing 0.05% Tween-20 (DPBS-T) for 1 h. The bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase (HRP) (Southern Biotech, cat. 2040-05, lot B3919-XD29, 1:5,000 dilution) and a 3,3′,5,5′-tetramethylbenzidine (TMB) substrate (Thermo Fisher Scientific). Colour development was monitored, 1M hydrochloric acid was added to stop the reaction, and the absorbance was measured at 450 nm using a spectrophotometer (Biotek). For dose–response assays, serial dilutions of purified monoclonal antibodies were applied to the wells in triplicate, and antibody binding was detected as detailed above. EC$_{50}$ values for binding were determined using Prism v.8.0 software (GraphPad) after log transformation of the monoclonal antibody concentration using sigmoidal dose–response nonlinear regression analysis.

RBD minimal human ACE2 recognition motif peptide binding ELISA

Wells of 384-well microtitre plates were coated with 1 μg ml^{-1} streptavidin at 4 °C overnight. Plates were blocked with 0.5% bovine serum albumin (BSA) in DPBS-T for 1 h. Plates were washed 4 times with 1× PBST and 2 μg ml^{-1} biotinylated ACE2 binding motif peptide (LT5578, from LifeTein, LLC) was added to bind streptavidin for 1 h at ambient temperature. Purified monoclonal antibodies were diluted in blocking buffer, added to the wells and incubated for 1 h at ambient temperature. The bound antibodies were detected using goat anti-human IgG conjugated with HRP (2014-05, Southern Biotech) and a TMB substrate (Thermo Fisher Scientific). Colour development was monitored, 1M hydrochloric acid was added to stop the reaction, and the absorbance was measured at 450 nm using a spectrophotometer (Biotek). For dose–response assays, serial threefold dilutions starting at a 10 μg ml^{-1} concentration of purified monoclonal antibodies were applied to the wells in triplicate, and antibody binding was detected as detailed above.

Analysis of binding of antibodies to variant RBD proteins with alanine or arginine point mutations

Biolayer light interferometry was performed using an Octet RED96 instrument (ForteBio; Pall Life Sciences) and wild-type RBD protein or a mutant RBD protein with a single amino acid change at defined positions to alanine or arginine. Binding of the RBD proteins was confirmed by first capturing 8×His-tagged RBD wild-type or mutant protein from a 10 μg ml^{-1} (around 200 nM) solution onto Penta-His biosensors for 300 s. The biosensor tips then were submerged in binding buffer (PBS/0.2% Tween 20) for a 60 s wash, followed by immersion in a solution containing 150 nM of monoclonal antibody for 180 s (association), followed by a subsequent immersion in binding buffer for 180 s (dissociation). The response for each RBD mutant protein was normalized to that of the wild-type RBD protein.

FRNT

Serial dilutions of monoclonal antibodies were incubated with 10^2 FFU of SARS-CoV-2 for 1 h at 37 °C. The antibody–virus complexes were added to Vero E6 cell-culture monolayers in 96-well plates for 1 h at 37 °C. Cells were then overlaid with 1% (w/v) methylcellulose in minimum essential medium (MEM) supplemented to contain 2% heat-inactivated FBS. Plates were fixed 30 h later by removing overlays and fixed with 4% paraformaldehyde (PFA) in PBS for 20 min at room temperature. The plates were incubated sequentially with 1 μg ml^{-1} of rCR3022 anti-S antibody[12] and HRP-conjugated goat anti-human IgG (Sigma-Aldrich, A6029) in PBS supplemented with 0.1% (w/v) saponin (Sigma) and 0.1% BSA. SARS-CoV-2-infected cell foci were visualized using TrueBlue peroxidase substrate (KPL) and quantitated on an ImmunoSpot 5.0.37 Macro Analyzer (Cellular Technologies). Data were processed using Prism v.8.0 (GraphPad). IC$_{50}$ values were determined by nonlinear regression analysis using Prism software.

Generation of S protein pseudotyped lentivirus

Suspension cultures of 293 cells were seeded and transfected with a third-generation HIV-based lentiviral vector expressing luciferase along with packaging plasmids encoding for the following: SARS-CoV-2 spike protein with a C-terminal 19 amino acid deletion, Rev, and Gag-pol. The medium was changed 16–20 h after transfection, and the supernatant containing virus was collected 24 h later. Cell debris was removed by low-speed centrifugation, and the supernatant was passed through a 0.45-μm filter unit. The pseudovirus was pelleted by ultracentrifugation and resuspended in PBS for a 100-fold concentrated stock.

Pseudovirus neutralization assay

Serial dilutions of monoclonal antibodies were prepared in a 384-well microtitre plate and pre-incubated with pseudovirus for 30 min at 37 °C, to which 293 cells that stably express human ACE2 were added. The plate was returned to the 37 °C incubator, and then 48 h later luciferase activity was measured on an EnVision 2105 Multimode Plate Reader (Perkin Elmer) using the Bright-Glo Luciferase Assay System (Promega), according to the manufacturer's recommendations. Per cent inhibition was calculated relative to pseudovirus-only control. IC$_{50}$ values were determined by nonlinear regression using Prism v.8.1.0 (GraphPad). The average IC$_{50}$ value for each antibody was determined from a minimum of three independent experiments.

Measurement of synergistic neutralization by a combination of antibodies

Synergy was defined as higher neutralizing activity mediated by a cocktail of two monoclonal antibodies when compared to that mediated by individual monoclonal antibodies at the same total concentration of antibodies in vitro. To assess whether two monoclonal antibodies synergize in a cocktail to neutralize SARS-CoV-2, we used a previously reported approach to quantify synergy[11]. To evaluate the significance of the beneficial effect from combining monoclonal antibodies, the observed combination responses (dose–response matrix) were compared with the expected responses calculated by means of synergy-scoring models[11]. Virus neutralization was measured in a conventional FRNT assay using wild-type SARS-CoV-2 and Vero E6 cell-culture monolayers. The individual monoclonal antibodies COV2-2196 and COV2-2130 were mixed at different concentrations to assess the neutralizing activity of different ratios of monoclonal antibodies in the cocktail. Specifically, each of seven two-fold dilutions of COV2-2130 (starting from 500 ng ml^{-1}) was mixed with each of the nine two-fold dilutions of COV2-2196 (starting from 500 ng ml^{-1}) in a total volume of 50 μl for each condition and then incubated with 50 μl of wild-type SARS-CoV-2 in cell culture medium (RPMI-1640 medium supplemented with 2% FBS) before applying to confluent Vero E6 cells grown in 96-well plates. The control values included those for determining the dose–response of the neutralizing activity measured separately for the individual monoclonal antibody COV2-2196 or COV2-2130, which were assessed at the same doses as in the cocktail. Each measurement was performed in duplicate. We next calculated the per cent virus neutralization for each condition and then calculated the synergy score value, which defines the interaction between these two monoclonal antibodies in the cocktail. A synergy score of less than −10 indicates antagonism, a score from −10 to 10 indicates an additive effect, and a score greater than 10 indicates a synergistic effect[28].

Quantification of monoclonal antibodies

Quantification of purified monoclonal antibodies was performed by UV spectrophotometry using a NanoDrop spectrophotometer and accounting for the extinction coefficient of human IgG.

Competition-binding analysis through biolayer interferometry

Anti-mouse IgG Fc capture biosensors (FortéBio 18-5089) on an Octet HTX biolayer interferometry instrument (FortéBio) were soaked for 10 min in 1× kinetics buffer (Molecular Devices 18-1105), followed by a baseline signal measurement for 60 s. Recombinant SARS-CoV-2 RBD fused to mouse IgG1 (RBD–mFc, Sino Biological 40592-V05H) was immobilized onto the biosensor tips for 180 s. After a wash step in 1× kinetics buffer for 30 s, the reference antibody (5 μg ml^{-1}) was incubated with the antigen-containing biosensor for 600 s. Reference antibodies included the SARS-CoV human monoclonal antibodies CR3022 and COV2-2196. After a wash step in 1× kinetics buffer for 30 s, the biosensor tips then were immersed into the second antibody (5 μg ml^{-1}) for 300 s. The maximum binding of each antibody was normalized to a buffer-only control. Self-to-self blocking was subtracted. A comparison between the maximum signal of each antibody was used to determine the per cent binding of each antibody. A reduction in maximum signal to less than 33% of the un-competed signal was considered full competition of binding for the second antibody in the presence of the reference antibody. A reduction in maximum signal to between 33% and 67% of the un-competed signal was considered intermediate competition of binding for the second antibody in the presence of the reference antibody. A per cent binding of the maximum signal of more than 67% was considered absence of competition of binding for the second antibody in the presence of the reference antibody.

Human ACE2 inhibition analysis

Wells of 384-well microtitre plates were coated with 1 μg ml^{-1} purified recombinant SARS-CoV-2 S2P$_{\mathrm{ecto}}$ protein at 4 °C overnight. Plates were blocked with 2% non-fat dry milk and 2% normal goat serum in DPBS-T for 1 h. For screening assays, purified monoclonal antibodies from microscale expression were diluted twofold in blocking buffer starting from 10 μg ml^{-1} in triplicate, added to the wells (20 μl per well) and incubated for 1 h at ambient temperature. Recombinant human ACE2 with a C-terminal Flag tag peptide was added to wells at 2 μg ml^{-1} in a 5 μl per well volume (final 0.4 μg ml^{-1} concentration of human ACE2) without washing of antibody and then incubated for 40 min at ambient temperature. Plates were washed and bound human ACE2 was detected using HRP-conjugated anti-Flag antibody (Sigma-Aldrich, cat. A8592, lot SLBV3799, 1:5,000 dilution) and TMB substrate. ACE2 binding without antibody served as a control. The signal obtained for binding of the human ACE2 in the presence of each dilution of tested antibody was expressed as a percentage of the human ACE2 binding without antibody after subtracting the background signal. For dose–response assays, serial dilutions of purified monoclonal antibodies were applied to the wells in triplicate, and monoclonal antibody binding was detected as detailed above. IC$_{50}$ values for inhibition by monoclonal antibody of S2P$_{\mathrm{ecto}}$ protein binding to human ACE2 was determined after log transformation of antibody concentration using sigmoidal dose–response nonlinear regression analysis (Prism v.8.0, GraphPad).

Human-ACE2-blocking assay using biolayer interferometry biosensor

Anti-mouse IgG biosensors on an Octet HTX biolayer interferometry instrument (FortéBio) were soaked for 10 min in 1× kinetics buffer, followed by a baseline signal measurement for 60 s. Recombinant SARS-CoV-2 RBD fused to mouse IgG1 (RBD–mFc, Sino Biological, 40592-V05H) was immobilized onto the biosensor tips for 180 s. After a wash step in 1× kinetics buffer for 30 s, the antibody (5 μg ml^{-1}) was incubated with the antigen-coated biosensor for 600 s. After a wash step in 1× kinetics buffer for 30 s, the biosensor tips then were immersed into the human ACE2 receptor (20 μg ml^{-1}) (Sigma-Aldrich, SAE0064) for 300 s. The maximum binding of human ACE2 was normalized to a buffer-only control. Per cent binding of human ACE2 in the presence of antibody was compared to human ACE2 maximum binding. A reduction in maximal signal to less than 30% was considered human-ACE2-blocking.

High-throughput competition-binding analysis

Wells of 384-well microtitre plates were coated with 1 μg ml^{-1} purified SARS-CoV-2 S2P$_{\mathrm{ecto}}$ protein at 4 °C overnight. Plates were blocked with 2% BSA in DPBS-T for 1 h. Microscale purified unlabelled monoclonal antibodies were diluted tenfold in blocking buffer, added to the wells (20 μl per well) in quadruplicates and incubated for 1 h at ambient temperature. A biotinylated preparation of a recombinant monoclonal antibody based on the variable gene sequence of the previously described monoclonal antibody CR3022[12], as well as the newly identified monoclonal antibodies COV2-2130 and COV2-2196 that recognized distinct antigenic regions of the SARS-CoV-2 S protein, were added to each of four wells with the respective monoclonal antibody at 2.5 μg ml^{-1} in a volume of 5 μl per well (final concentration of biotinylated monoclonal antibody 0.5 μg ml^{-1}) without washing of unlabelled antibody, and then incubated for 1 h at ambient temperature. Plates were washed and bound antibodies were detected using HRP-conjugated avidin (Sigma) and a TMB substrate. The signal obtained for binding of the biotin-labelled reference antibody in the presence of the unlabelled tested antibody was expressed as a percentage of the binding of the reference antibody alone after subtracting the background signal. Tested monoclonal antibodies were considered competing if their presence reduced the reference antibody binding to less than 41% of its maximal binding and non-competing if the signal was greater than 71%. A level of 40–70% was considered intermediate competition.

Plasma or serum antibody competition-binding assays

Wells of 384-well microtitre plates were coated with 1 μg ml^{-1} purified SARS-CoV-2 S2P$_{\mathrm{ecto}}$ at 4 °C overnight. Plates were blocked with 2% BSA in DPBS-T for 1 h. Plasma or serum samples were diluted in blocking buffer twofold starting from 1:10 sample dilution, added to the wells (20 μl per well) in triplicate and incubated for 1 h at ambient temperature. For self-blocking controls, unlabelled monoclonal antibodies COV2-2196 or COV2-2130 were added at 10 μg ml^{-1} to separate wells coated with S2P$_{ecto}$. Serum from a donor without an exposure history to SARS-CoV-2 was used as a negative control for monoclonal antibody binding inhibition. A biotinylated monoclonal antibody COV2-2196 or COV2-2130 was added to the respective wells at 2.5 μg ml^{-1} in a volume of 5 μl per well (final concentration of biotinylated monoclonal antibody 0.5 μg ml^{-1}) without washing of unlabelled antibody, and then incubated for 30 min at ambient temperature. Binding of biotinylated monoclonal antibodies COV2-2196 or COV2-2130 alone to S2P$_{ecto}$ served as a control for maximum binding. Plates were washed and bound antibodies were detected using HRP-conjugated avidin (Sigma) and a TMB substrate. Inhibition of COV2-2196 or COV2-2130 binding in the presence of each dilution of tested plasma or serum was calculated as a percentage of the maximum COV2-2196 or COV2-2130 binding inhibition using values from COV2-2196 or COV2-2130 binding alone (maximum binding) and the corresponding self-blocking controls (maximum inhibition) after subtracting the background signal. For the human ACE2 inhibition assay by plasma or serum antibodies, plasma or serum samples were diluted and added to wells with S2P$_{ecto}$ as detailed above. Recombinant human ACE2 was added to wells at 2 μg ml^{-1}in a volume of 5 μl per well (final concentration of human ACE2 0.4 μg ml^{-1}) without washing of antibody, and then incubated for 40 min at ambient temperature. Plates were washed and bound human ACE2 was detected using HRP-conjugated anti-Flag antibody (Sigma) and a TMB substrate. Human ACE2 binding without antibody served as a control. The signal obtained for binding of the human ACE2 in the presence of each dilution of tested plasma or serum was expressed as a percentage of the ACE2 binding without antibody after subtracting the background signal.

Protection against wild-type SARS-CoV-2 in mice transduced with human ACE2

Animal studies were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Institutional Animal Care and Use Committee at the Washington University School of Medicine (assurance number A3381–01). Viral inoculations were performed under anaesthesia, which was induced and maintained with ketamine hydrochloride and xylazine, and all efforts were made to minimize animal suffering.

Wild-type, female BALB/c mice were purchased from The Jackson Laboratory (strain 000651). Mice were housed in groups of up to 5 mice per cage at 18–24 °C ambient temperatures and 40–60% humidity. Mice were fed a 20% protein diet (PicoLab 5053, Purina) and maintained on a 12-h light–dark cycle (06:00 to 18:00). Food and water were available ad libitum.

Mice (10–11 weeks old) were given a single intraperitoneal injection of 2 mg of anti-IFNAR1 monoclonal antibody (MAR1-5A3[55], Leinco) one day before intranasal administration of 2.5×10^8 PFU of AdV-hACE2. Five days after AdV transduction, mice were inoculated with 4×10^5 PFU of SARS-CoV-2 via the intranasal route. Anti-SARS-CoV-2 human monoclonal antibodies or isotype control monoclonal antibodies were administered 24 h before (prophylaxis) or 12 h after (therapy) SARS-CoV-2 inoculation. Weights were monitored on a daily basis, mice were euthanized at 2 or 7 dpi and tissues were collected.

Measurement of viral burden

For RT–qPCR, tissues were weighed and homogenized with zirconia beads in a MagNA Lyser instrument (Roche Life Science) in 1 ml of DMEM medium supplemented with 2% heat-inactivated FBS. Tissue homogenates were clarified by centrifugation at 10,000 rpm for 5 min and stored at −80 °C. RNA was extracted using a MagMax mirVana Total RNA isolation kit (Thermo Fisher Scientific) and a Kingfisher Flex 96-well extraction machine (Thermo Fisher Scientific). TaqMan primers were designed to target a conserved region of the *N* gene using SARS-CoV-2 (MN908947) sequence as a guide (L primer: ATGCTGCAATCGTGCT ACAA; R primer: GACTGCCGCCTCTGCTC; probe: /56-FAM/TCA AGGAAC/ZEN/AACATTGCCAA/3IABkFQ/). To establish an RNA standard curve, we generated concatenated segments of the *N* gene in a gBlocks fragment (IDT) and cloned this into the PCR-II topo vector (Invitrogen). The vector was linearized, and in vitro T7-DNA-dependent RNA transcription was performed to generate materials for a quantitative standard curve.

For the plaque assay, homogenates were diluted serially tenfold and applied to Vero-furin cell monolayers in 12-well plates. Plates were incubated at 37 °C for 1 h with rocking every 15 min. Cells were then overlaid with 1% (w/v) methylcellulose in MEM supplemented with 2% FBS. Plates were collected 72 h later by removing overlays and fixed with 4% PFA in PBS for 20 min at ambient temperature. After removing the 4% PFA, plaques were visualized by adding 1 ml per well 0.05% crystal violet in 20% methanol for 20 min at ambient temperature. Excess crystal violet was washed away with PBS, and plaques were counted.

Cytokine and chemokine mRNA measurements

RNA was isolated from lung homogenates at 7 dpi as described above. cDNA was synthesized from DNase-treated RNA using the High-Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific) with the addition of RNase inhibitor, following the manufacturer's protocol. Cytokine and chemokine expression was determined using TaqMan Fast Universal PCR master mix (Thermo Fisher Scientific) with commercial primers and probe sets specific for *Ifng* (IDT: Mm.PT.58.41769240), *Il6* (Mm.PT.58.10005566), *Cxcl10* (Mm. PT.58.43575827) and *Ccl2* (Mm.PT.58.42151692) and results were normalized to *Gapdh* (Mm.PT.39a.1) levels. Fold change was determined using the $2^{-\Delta\Delta Ct}$ method comparing anti-SARS-CoV-2-specific or isotype-control monoclonal-antibody-treated mice to naive controls.

Histology

Mice were euthanized, and tissues were collected before lung inflation and fixation. The left lung lobe was tied off at the left main bronchus and collected for viral RNA analysis. The right lung lobe was inflated with around 1.2 ml of 10% neutral buffered formalin using a 3-ml syringe and catheter inserted into the trachea. For fixation after infection, inflated lungs were kept in a 40-ml suspension of neutral buffered formalin for 7 days before further processing. Tissues were embedded in paraffin, and sections were stained with haematoxylin and eosin. Tissue sections were visualized using a Leica DM6B microscope equipped with a Leica DFC7000T camera. The sections were scored by an immunopathology expert blinded to the compositions of the groups.

Viral challenge studies using MA-SARS-CoV-2 and wild-type mice

Animal studies were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Institutional Animal Care and Use Committee at the UNC Chapel Hill School of Medicine (NIH/PHS animal welfare assurance number D16-00256 (A3410-01)). Virus inoculations were performed under anaesthesia that was induced and maintained with ketamine hydrochloride and xylazine, and all efforts were made to minimize animal suffering.

Protection against MA-SARS-CoV-2 in wild-type mice

BALB/c mice (12 months old) from Envigo were used in experiments. Mice were housed in groups of up to 5 mice per cage at 18–24 °C ambient temperatures and 40–60% humidity. Mice were fed a 20% protein diet (PicoLab 5053, Purina) and maintained on a 12-h light–dark cycle (08:00 to 20:00). Food and water were available ad libitum. Mice were acclimated in the BSL3 for at least 72 h before start of experiments. At

6 h before infection, mice were treated with 200 µg of human monoclonal antibodies via intraperitoneal injection. The next day, mice were anaesthetized with a mixture of ketamine and xylazine and intranasally inoculated with 10^5 PFU of MA-SARS-CoV-2 diluted in PBS. Daily weight loss was measured, and at 2 dpi mice were euthanized by isoflurane overdose before tissue collection. For the post-exposure therapy study, mice were inoculated intranasally with 10^5 PFU of MA-SARS-CoV-2 and 12 h later given the indicated antibody treatments by intraperitoneal injection. The lungs were collected at 2 dpi.

Plaque assay of lung tissue homogenates

The lower lobe of the right lung was homogenized in 1 ml PBS using a MagnaLyser (Roche). Serial dilutions of virus were titrated on Vero E6 cell-culture monolayers, and virus plaques were visualized by neutral red staining two days after inoculation. The limit of detection for the assay is 100 PFU per lung.

NHP challenge study

The NHP research studies adhered to principles stated in the eighth edition of the Guide for the Care and Use of Laboratory Animals. The facility in which this research was conducted (Bioqual, Rockville) is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC) and approved by the Office of Laboratory Animal Welfare (NIH/PHS assurance number D16-00052). NHP studies were conducted in compliance with all relevant local, state and federal regulations and were approved by the Institutional Animal Care and Use Committee (IACUC) at Bioqual.

Twelve healthy adult rhesus macaques (*Macaca mulatta*) of Indian origin (5–15 kg body weight) were studied. Rhesus macaques were 5–7 years old and mixed male and female. Macaques were allocated randomly to two anti-SARS-CoV-2 monoclonal antibody treatment groups ($n = 4$ per group) and one control (isotype-treated) group ($n = 4$ per group). Macaques received one 50 $\mathrm{mg\ kg^{-1}}$ dose of COV2-2196, COV2-2381 or an isotype control monoclonal antibody intravenously on day −3 and were challenged three days later with 1.1×10^4 PFU SARS-CoV-2, administered as 1 ml via the intranasal route and 1 ml via the intratracheal route. After challenge, viral RNA was assessed by RT–qPCR in bronchoalveolar lavage and nasal swabs at multiple time points as described previously[34,35]. All macaques were given physical examinations. In addition, all macaques were monitored daily with an internal scoring protocol approved by the IACUC. These studies were not blinded.

Detection of circulating human monoclonal antibodies in NHP serum

ELISA plates were coated overnight at 4 °C with 1 $\mathrm{\mu g\ ml^{-1}}$ of goat anti-human IgG (H+L) secondary antibody (monkey pre-adsorbed) (Novus Biologicals, NB7487) and then blocked for 2 h. The serum samples were assayed at threefold dilutions starting at a 1:3 dilution in Blocker Casein in PBS (Thermo Fisher Scientific) diluent. Samples were incubated for 1 h at ambient temperature and then removed, and plates were washed. Wells then were incubated for 1 h with HRP-conjugated goat anti-human IgG (monkey pre-adsorbed) (Southern Biotech, 2049-05) at a 1:4,000 dilution. Wells were washed and then incubated with SureBlue Reserve TMB Microwell Peroxidase Substrate (Seracare) (100 µl per well) for 3 min followed by TMB Stop Solution (Seracare) to stop the reaction (100 µl per well). Microplates were read at 450 nm. The concentrations of the human monoclonal antibodies were interpolated from the linear range of purified human IgG (Sigma) standard curves using Prism v.8.0 (GraphPad).

Quantification and statistical analysis

Mean ± s.e.m. or mean ± s.d. were determined for continuous variables as noted. Technical and biological replicates are described in the figure legends. In the mouse studies, the comparison of weight-change curves was performed using a repeated measurements two-way ANOVA with Tukey's post hoc test using Prism v.8.0 (GraphPad). Viral burden and gene-expression measurements were compared using a Kruskal–Wallis ANOVA with Dunn's post hoc test or a two-sided Mann–Whitney *U*-test using Prism v.8.0 (GraphPad). The analyses of synergy score and the dose–response matrix were performed using a web application, SynergyFinder[28].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The electron microscopy maps have been deposited at the Electron Microscopy Data Bank (EMDB) with accession codes EMD-21974, EMD-21975, EMD-21976 and EMD-21977 (Supplementary Table 2). The electron microscopy map EMD-21965 is publicly available. The accession numbers for the cryo-electron-microscopy and crystal structures used for structural analysis, including structures of the closed conformation of SARS-CoV-2 S (PDB: 6VXX), the open conformation of SARS-CoV-2 (PDB: 6VYB), the Fab used for docking (PDB: 12E8) and the SARS-CoV-2 RBD–human ACE2 complex (PDB: 6M0J) are publicly available. Sequences of the monoclonal antibodies characterized here are available from GenBank under the following accession numbers: MT665032–MT665070, MT665419–MT665457, MT763531 and MT763532. Materials used in this study will be made available but may require execution of a Materials Transfer Agreement. Source data are provided with this paper.

47. Mukherjee, S. et al. Enhancing dengue virus maturation using a stable furin over-expressing cell line. *Virology* **497**, 33–40 (2016).
48. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification – powerful tools in modern electron microscopy. *Biol. Proced. Online* **6**, 23–34 (2004).
49. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
50. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
51. Bepler, T., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM. Preprint at bioRxiv https://doi.org/10.1101/838920 (2019).
52. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Chng, J. et al. Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *MAbs* **7**, 403–412 (2015).
55. Sheehan, K. C. et al. Blocking monoclonal antibodies specific for mouse IFN-α/β receptor subunit 1 (IFNAR-1) from mice immunized by in vivo hydrodynamic transfection. *J. Interferon Cytokine Res.* **26**, 804–819 (2006).

Acknowledgements We thank A. Jones and the staff of the Vanderbilt Technologies for Advanced Genomics (VANTAGE) core laboratory for expedited sequencing; R. Trosseth for assistance with data management and analysis; R. Bombardi and C. Soto of VUMC for technical consultation on genomics approaches; A. Kim, A. Bailey, L. VanBlargan and J. Earnest of WUSTL for experimental assistance and key reagents; K. M. Tuffy, S. Diallo, P. M. McTamney and L. Clarke of AstraZeneca for the generation of protein and pseudovirus reagents and related data; and H. Andersen, M. G. Lewis, R. Nityanandam, M. Kirilova and K. Verrington for research assistance with the NHP studies. This study was supported by Defense Advanced Research Projects Agency (DARPA) grants HR0011-18-2-0001 and HR00 11-18-3-0001; NIH contracts 75N93019C00074 and 75N93019C00062; NIH grants U01 AI150739, R01 AI130591 and R35 HL145242; the Dolly Parton COVID-19 Research Fund at Vanderbilt; and NIH grant S10 RR028106 for the Next Generation Nucleic Acid Sequencer, housed in VANTAGE and the Vanderbilt Institute for Clinical and Translational Research with grant support from UL1TR002243 from NCATS/NIH. S.J.Z. was supported by NIH T32 AI095202; J.B.C. was supported by a Helen Hay Whitney Foundation postdoctoral fellowship; B.T.M. was supported by NIH F32 AI138392; D.R.M. was supported by NIH T32 AI007151 and a Burroughs Wellcome Fund Postdoctoral Enrichment Program Award; L.E.W. was supported by NIH F31 AI145189; E.C.C. was supported by NIH T32 AI138932; and J.E.C. is the recipient of the 2019 Future Insight Prize from Merck KGaA, which supported this research with a research grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Author contributions S.J.Z., P.G., R.H.C., L.B.T., M.S.D. and J.E.C. conceived the project; J.E.C. and M.S.D. obtained funding; S.J.Z., P.G., J.B.C., E.B., R.E.C., J.P.N., A.S., J.X.R., A.T., R.S.N., R.E.S., N.S., D.R.M., L.E.W., A.O.H., N.M.K., E.S.W., J.M.F., S.S., B.K.M., A.C., N.B.M., J.J.S., K.R., Y.-M.L., S.P.K., M.J.H., L.E.G. and L.B.T. performed laboratory experiments; E.C.C., T.J., S.D., L.M. and

B.T.M. performed computational work; J.M., N.L.K., D.H.B., R.S.B., L.B.T., M.S.D., R.H.C. and J.E.C. supervised the research; and S.J.Z., P.G., R.H.C. and J.E.C. wrote the first draft of the paper. All authors reviewed and approved the final manuscript.

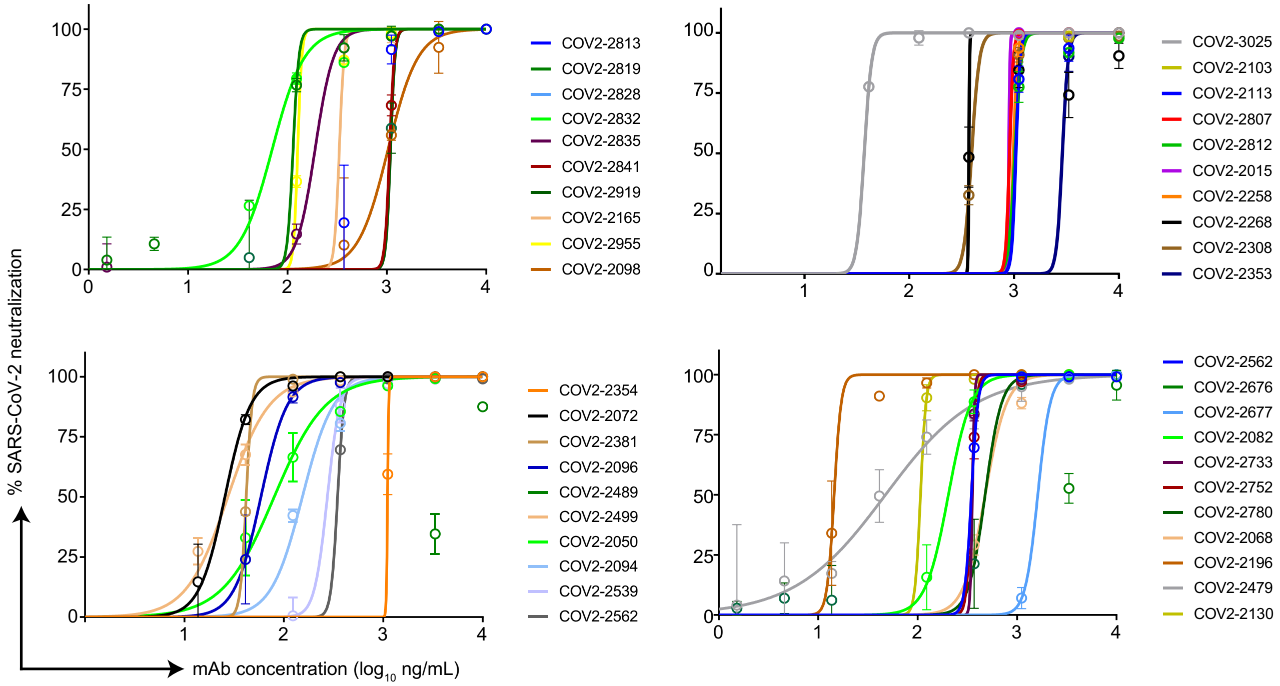
Competing interests R.S.B. has served as a consultant for Takeda and Sanofi Pasteur on issues related to vaccines. M.S.D. is a consultant for Inbios, Vir Biotechnology, NGM Biopharmaceuticals and Eli Lilly; is on the Scientific Advisory Board of Moderna; is a past recipient of an unrelated research grant from Moderna; and is a current recipient of an unrelated research grant from Emergent BioSolutions. J.E.C. has served as a consultant for Sanofi; is on the Scientific Advisory Boards of CompuVax and Meissa Vaccines; is a recipient of previous unrelated research grants from Moderna and Sanofi; and is a founder of IDBiologics. Vanderbilt University has applied for patents concerning SARS-CoV-2 antibodies that are related to this work. AstraZeneca has filed patents for materials and findings that are related to this work. J.J.S., K.R., Y.-M.L. and N.L.K. are employees of AstraZeneca and currently hold AstraZeneca stock or stock options. M.J.H. is a member of a data safety monitoring board for AstraZeneca and a founder of NuPeak Therapeutics. All other authors declare no competing interests.

Additional information

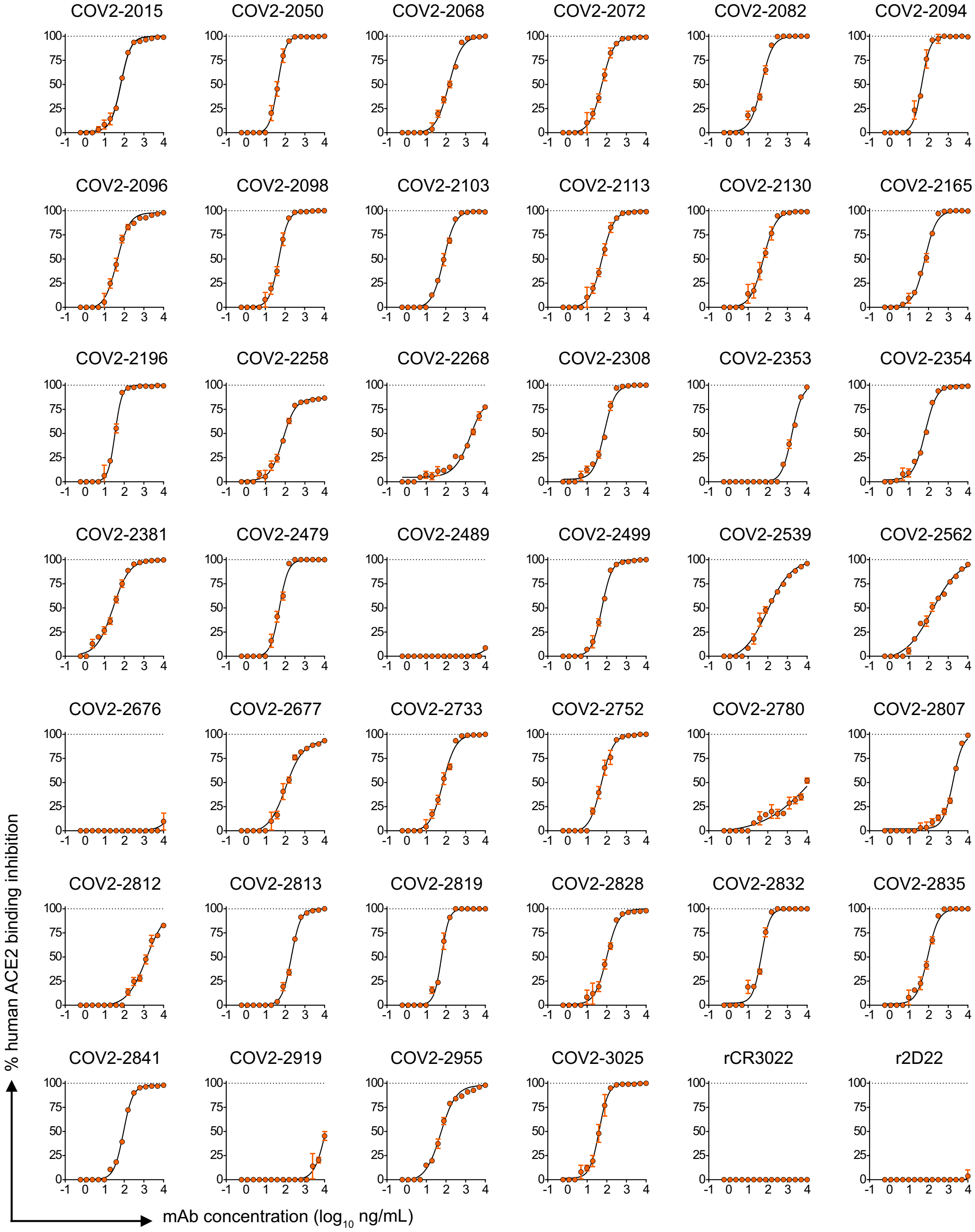
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2548-6.

Correspondence and requests for materials should be addressed to R.H.C. or J.E.C.

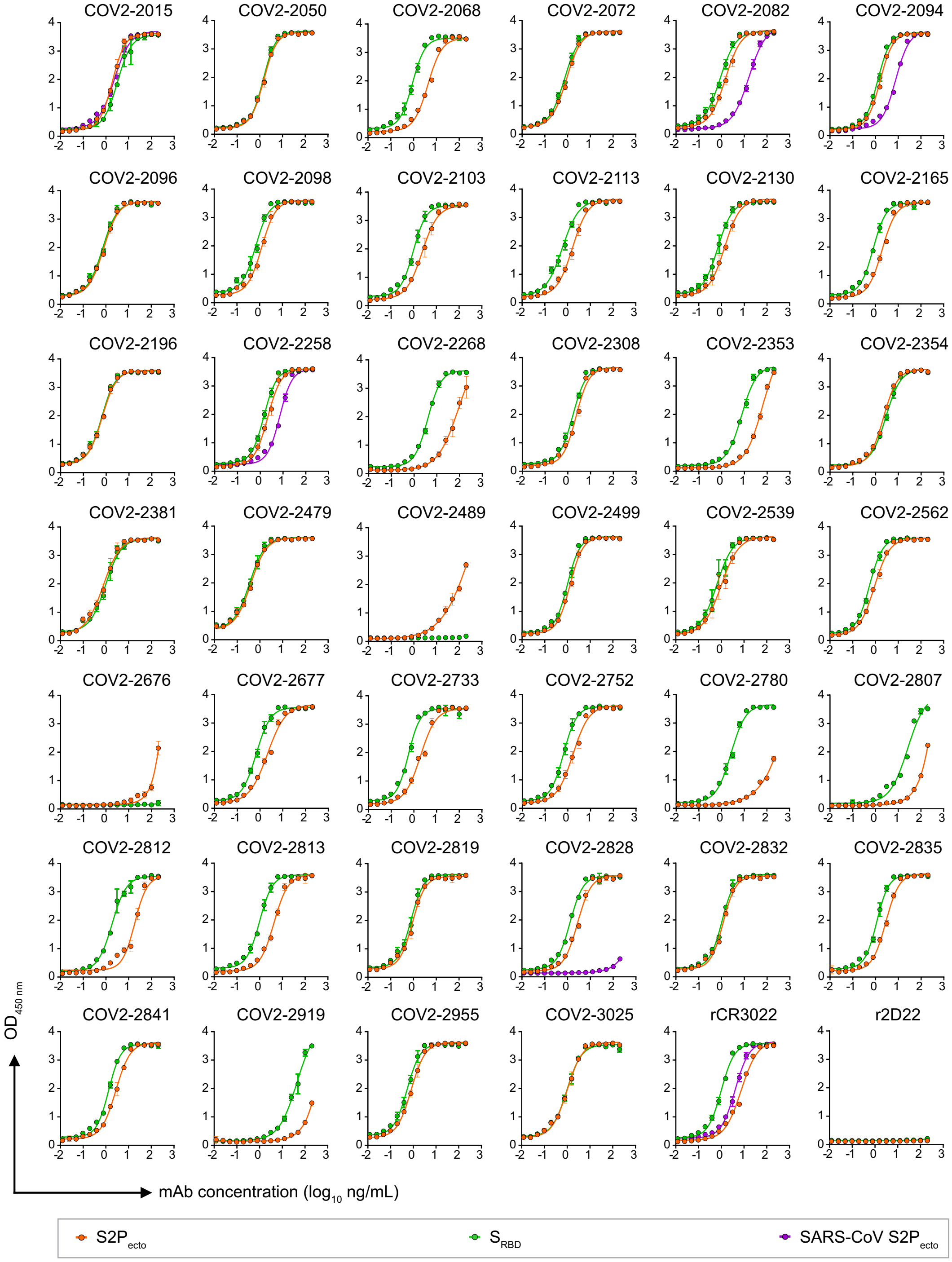
Peer review information *Nature* thanks Linda Saif and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



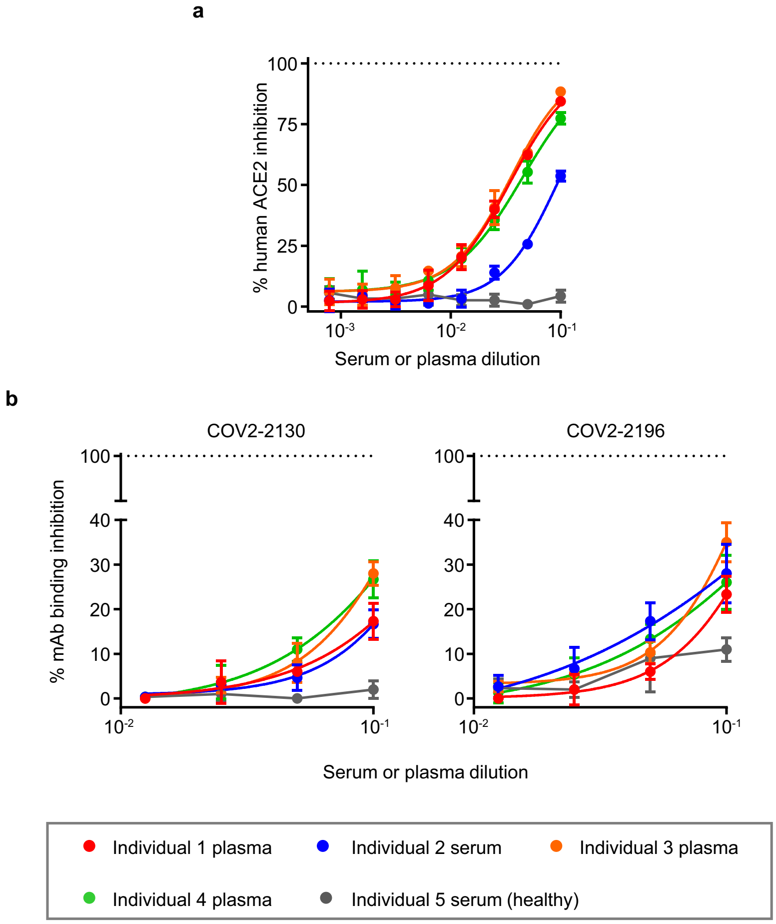
Extended Data Fig. 1 | SARS-CoV-2 neutralization curves for monoclonal antibody panel. Neutralization of wild-type SARS-CoV-2 by human monoclonal antibodies. Data are mean ± s.d. of technical duplicates, and represent one of two or more independent experiments.



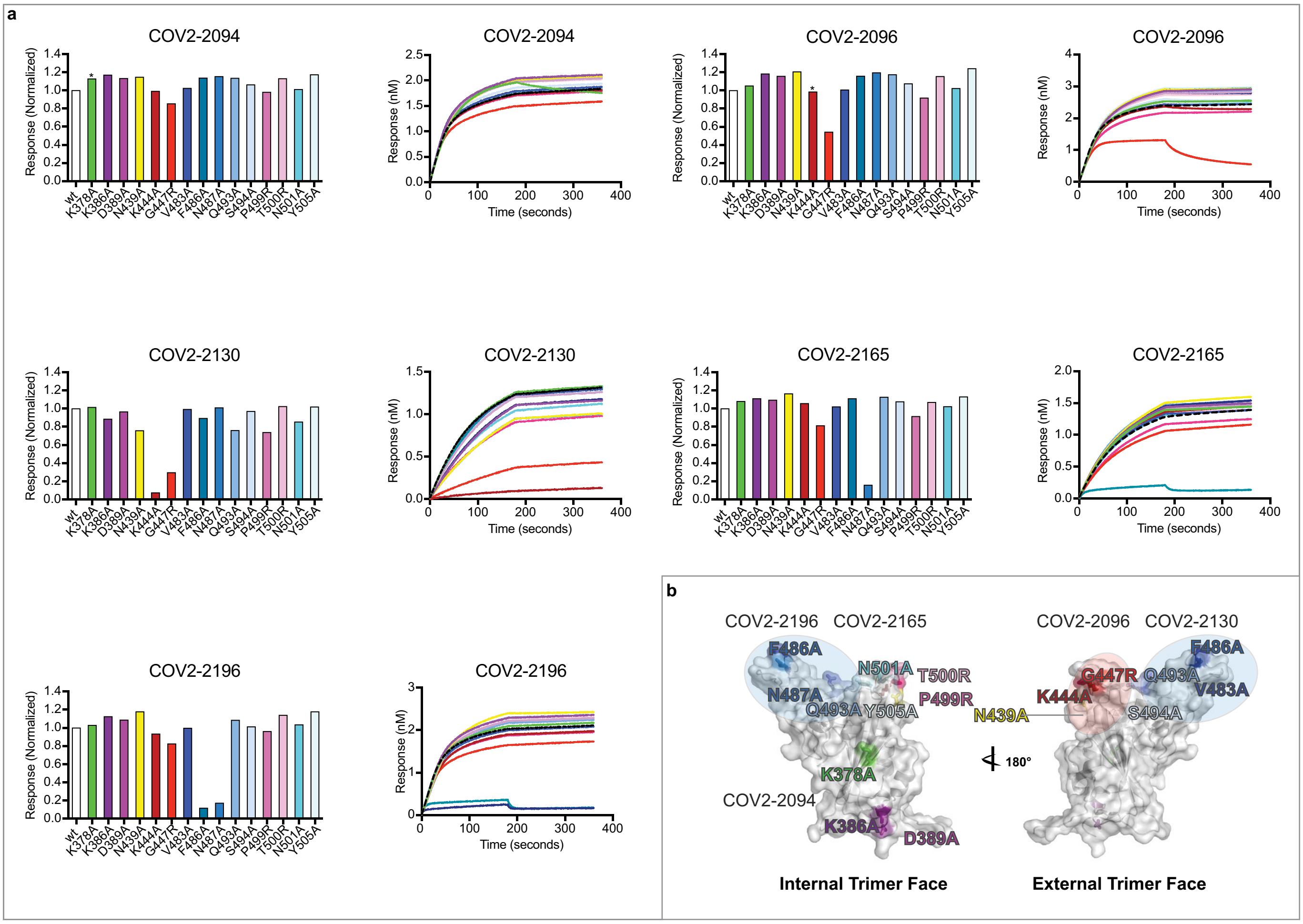
Extended Data Fig. 2 | Inhibition curves for monoclonal antibody inhibition of $\mathrm{S2P}_{\mathrm{ecto}}$ binding to human ACE2. Blocking of human ACE2 binding to $\mathrm{S2P}_{\mathrm{ecto}}$ by anti-SARS-CoV-2 neutralizing human monoclonal antibodies. Data are mean ± s.d. of triplicates of one experiment. Antibodies rCR3022 and r2D22 served as controls.



Extended Data Fig. 3 | ELISA binding of anti-SARS-CoV-2 neutralizing human monoclonal antibodies to trimeric S_{RBD}, $S2P_{ecto}$ or SARS-CoV $S2P_{ecto}$ antigen. Data are mean ± s.d. of triplicates, and are representative of two experiments. Antibodies rCR3022 and r2D22 served as controls.



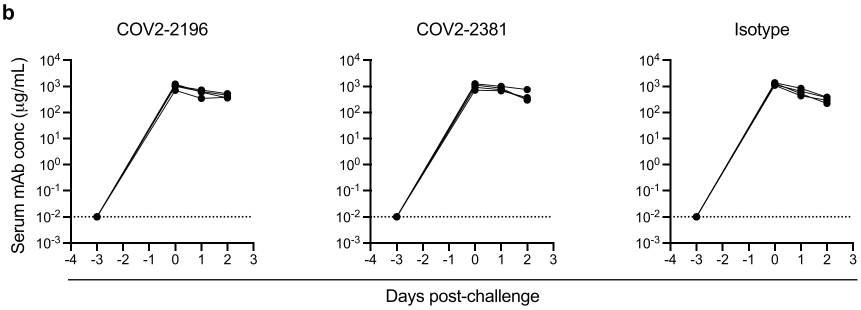
Extended Data Fig. 4 | Competition-binding analysis of serum or plasma antibodies with human ACE2 and monoclonal antibodies. a, Inhibition of human ACE2 binding to S2P_{ecto} by serum or plasma of four SARS-CoV-2 immune individuals or one non-immune control individual in an ELISA using SARS-CoV-2 S2P_{ecto}. Monoclonal antibodies were isolated from individuals 3 and 4 as described previously[5]. Data are mean ± s.d. of triplicates of one experiment. Dotted line indicates full inhibition (100%) of human ACE2 by 500 ng ml^{-1} of monoclonal antibody COV2-2196 or COV2-2130 that were used as controls for full human ACE2 inhibition. **b**, Inhibition of monoclonal antibody COV2-2130 (left) or COV2-2196 (right) binding to S2P_{ecto} by serum or plasma of four SARS-CoV-2 immune individual or one non-immune control individual in an ELISA using SARS-CoV-2 S2P_{ecto}. Data are mean ± s.d. of triplicates, and are representative of two experiments. Dotted line indicates the percentage of self-competition of monoclonal antibodies COV2-2196 and -2130 on the SARS-CoV-2 S2P_{ecto} antigen.



Extended Data Fig. 5 | Mapping of critical contact residues for monoclonal antibodies by alanine and arginine mutagenesis and biolayer interferometry. **a**, Bar graphs show response values for monoclonal antibody binding to wild-type or mutant S_{RBD} constructs normalized to the wild type. Asterisks denote residues where increased dissociation of monoclonal antibody was observed, probably indicating that the residue is proximal to the monoclonal antibody epitope. Full response curves for monoclonal antibody association and dissociation with wild-type or mutant S_{RBD} constructs are also shown. **b**, Structure of the RBD, highlighting the critical contact residues for several monoclonal antibodies and their location on the structure.

a

	Heavy chain features				Light chain features		
mAb	V Gene	D Gene	J Gene	HCDR3	V Gene	J Gene	LCDR3
COV2-2130	IGHV3-15	IGHD3-22	IGHJ4	`TTAGSYYYDTVGPGLPEGKFDY`	IGKV4-1	IGKJ4	`QQYYSTLT`
COV2-2196	IGHV1-58	IGHD2-2	IGHJ3	`AAPYCSSISCNDGFDI`	IGKV3-20	IGKJ1	`QHYGSSRGWT`
COV2-2381	IGHV1-58	IGHD2-2	IGHJ3	`AAPYCSRTSCHDAFDI`	IGKV3-20	IGKJ1	`QHFGSSSQWT`



Extended Data Fig. 6 | Sequence features of the human monoclonal antibodies used in animal studies and monoclonal antibody pharmacokinetics following their administration to NHPs. a, Sequence features of human monoclonal antibodies tested in animal models. Inferred variable genes are indicated and CDR3 amino acids are shown for heavy and light chains. **b**, Macaques received one 50 mg kg^{-1} dose of COV2-2196, COV2-2381 or an isotype control monoclonal antibody (*n* = 4 macaques per group) intravenously on day −3 and then were challenged intranasally and intratracheally with SARS-CoV-2 at day 0. The concentration of human monoclonal antibodies was determined at indicated time points. Each curve shows an individual macaque. Data represent a single experiment.

nature research

Corresponding author(s): James E. Crowe

Last updated by author(s): 8/01/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The microscope was operated using SerialEM software version 3.7 (PMID: 16182563). Negative stain electron microscopy image acquisition and processing was performed using the cryoSPARC software package version 2.14.2 (PMID: 28165473). The images were denoised and picked with Topaz software version 0.2.3 (bioRxiv. doi:10.1101/838920).
Data analysis	This study used commercially available GraphPad Prism software v8 for data representation and statistical analysis (GraphPad Prism; RRID: SCR_002798). Synergy was estimated using open source SynergyFinder software https://synergyfinder.fimm.fi/ (PMID: 28379339). UCSF chimera was used for molecular docking to the electron microscopy maps (UCSF Chimera; RRID: SCR_004097). Pymol was used to visualize molecular structures and freely available from https://www.pymol.org/2//.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The electron microscopy maps have been deposited at the Electron Microscopy Data Bank (EMDB) with accession codes EMD-21974, EMD-21975, EMD-21976 and EMD-21977 (Supplementary Table 2). The electron microscopy map EMD-21965 is publicly available. The acces- sion numbers for the cryo-electron-microscopy and crystal structures used for structural analysis, including structures of the closed con- formation of SARS-CoV-2 S (PDB: 6VXX), the open conformation of SARS-CoV-2 (PDB: 6VYB), the Fab used for docking (PDB: 12E8) and the SARS-CoV-2 RBD–human ACE2 complex (PDB: 6M0J) are publicly available. Sequences of the monoclonal antibodies characterized here are available from GenBank under the following accession num- bers: MT665032–MT665070, MT665419–MT665457, MT763531 and

MT763532. Materials used in this study will be made available but may require execution of a Materials Transfer Agreement. Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculations were performed to power each study. Sample sizes for mouse studies were determined based on our previous results for similar in vivo experiments which showed that the use of 5—10 mice per group represents a minimally sufficient sample size to produce a study power of >80% (adequacy standard used in most research). To ascertain reproducibility, studies for key experimental findings that include in vivo protection in mice by identified neutralizing mAbs were confirmed using two different mouse challenge models, and in prophylaxis and therapy settings with sample sizes of n=8-10 animals per experiment. Details about research subjects groups are provided in Supplementary information. Details about groups and sample sizes for mouse virus challenge studies are provided in the Results section and figure legends. For the NHP study, sample sizes were sufficient given large differences in viral load between treated and isotype control groups. The other key experiments that included in vitro measurements of antibody binding, hACE2 blocking, and virus neutralizing activities were carried out with two or more independent study replicates, which were sufficient given the large difference between activities for identified SARS-CoV-2-specific mAbs and isotype controls.
Data exclusions	No data were excluded from the analysis
Replication	Studies that were repeated are noted in figure captions and include all studies that demonstrated the key results reported in the manuscript. No studies have been reported that failed upon repetition. Antibodies of known activity were included across all experiments to verify reproducibility (e.g. presence of binding, blocking, or neutralizing activities), and included comparisons of newly identified SARS-CoV-2-specific antibodies to relevant characterized antibodies (e.g. rCR3022) and isotype matched antibody controls. These controls were included in each replicate experiment that measured binding, blocking, neutralizing, and in vivo protective activity of characterized anti-SARS-CoV-2 mAbs. Consistency of mAb activity across in vitro and in vivo experiments within this study indicate a high level of reproducibility.
Randomization	Animals were randomly allocated to the groups. For experiments other than animal studies, randomization is not relevant as this is an observational study.
Blinding	The investigators were not blinded for most studies except lung pathology evaluation. We used conventional antigen binding and virus neutralization assays using actual binding to the SARS-CoV2 spike antigen and live SARS-CoV2 neutralization as the readouts. We used two different challenge models to measure protective capacity of identified mAbs. In the first more stringent challenge model, we monitored for protection against severe weight loss using body weight measurement as a readout and RT-qPCR to quantify viral burden. In the second less stringent model using a mouse-adapted virus, we measured viral load using RT-qPCR and plaque assay for the infectious virus. For lung study pathology, H&E stained tissue sections were scored by an experienced immunopathologist blinded to the compositions of the groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	In a newly developed SARS-CoV-2 infection model in BALB/c mice in which human ACE2 is expressed in the lung after intranasal adenovirus (AdV-hACE2) transduction, mice were treated with anti-Ifnar1 mAb (MAR1-5A3; Leinco). Polyclonal goat anti-human IgG-HRP antibody (Southern Biotech Cat# 2040-05, Lot B3919-XD29) was used for antigen binding ELISA assays. Monoclonal anti-FLAG M2-Peroxidase (HRP) antibody produced in mouse (Sigma-Aldrich Cat# A8592, Lot SLBV3799) was used as a detection

antibody for hACE2 binding assays. For FRNT assay, previously described human anti-SARS-CoV rCR3022 antibody (PMID: 32245784) was used as a primary antibody and the detection was performed using anti-human IgG (γ-chain specific)-peroxidase antibody produced in goat (Sigma-Aldrich Cat# A6029). Capture antibody used for human mAb detection in NHP serum utilized a goat anti-human IgG (H+L) secondary antibody (monkey pre-adsorbed) (Novus Biological Cat# NB7487). Detection antibody used for human mAb detection in NHP serum utilized an HRP-labeled goat anti-human IgG (H+L), (monkey pre-adsorbed) (Southern Biotech Cat# 2049-05).
Newly discovered SARS-CoV2 spike antigen-specific monoclonal antibodies are described in this paper.

Validation

Newly discovered SARS-CoV2 spike antigen-specific monoclonal antibodies were validated via antigen binding, virus neutralization, and in vivo protection studies described in this paper. Validation of anti-Ifnar1 mAb activity was previously described (PMID: 17115899). Validation of rCR3022 antibody activity was previously described (PMID: 32245784). All other antibodies are commercially available. Antibodies used in a specific species or application have been appropriately validated by manufacturers and this information is provided on their website and information datasheets as follows:
Goat anti-human IgG-HRP (https://www.southernbiotech.com/?catno=2040-05&type=Polyclonal#&panel1-1&panel2-1);
Anti-human IgG (γ-chain specific)-peroxidase antibody produced in goat (https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/6/a6029dat.pdf);
Monoclonal anti-FLAG M2-Peroxidase (HRP) antibody produced in mouse (https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/6/a8592dat.pdf);
Goat anti-human IgG (H+L) secondary antibody (monkey pre-adsorbed) (https://www.novusbio.com/PDFs/NB7487.pdf);
Goat anti-human IgG, monkey ads-HRP (https://www.southernbiotech.com/techbul/2049.pdf).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

In this study we used the following cell lines: Vero E6 (American Type Culture Collection (ATCC), Cat# CRL-1586), Vero (ATCC Cat# CCL-81), HEK293 (ATCC Cat# CRL-1573), and HEK293T (ATCC Cat# CRL-3216), Expi293F (ThermoFisher Scientific, A1452), FreeStyle 293-F (ThermoFisher Scientific, R79007), and ExpiCHO (ThermoFisher Scientific, A29127). Vero-furin cells were obtained from T. Pierson (NIH) and have been previously described (PMID: 27420797).

Authentication

None of the cell lines used were authenticated

Mycoplasma contamination

All cell lines were tested and confirmed negative for Mycoplasma contamination

Commonly misidentified lines
(See ICLAC register)

None

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

For viral challenge using authentic SARS-CoV-2, wild-type female BALB/c mice (10-11-week-old) that were purchased from Jackson Laboratory (strain 000651) were used. Animals were housed in groups of up to 5 mice/cage at 18-24°C ambient temperatures and 40-60% humidity. Mice were fed a 20% protein diet (PicoLab 5053, Purina) and maintained on a 12 hour light/dark cycle (6 am to 6 pm). Food and water were available ad libitum.

For viral challenge using MA-SARS-CoV-2, wild-type female 12-month-old BALB/c mice from Envigo (strain 047) were used. Animals were housed in groups of up to 5 mice/cage at 18-24°C ambient temperatures and 40-60% humidity. Mice were fed a 20% protein diet (PicoLab 5053, Purina) and maintained on a 12 hrs light/dark cycle (8 am to 8 pm). Food and water were available ad libitum.

Twelve healthy adult rhesus macaques (Macaca mulatta) of Indian origin (5 to 15 kg body weight) were studied. Rhesus macaques were 5-7 years old and mixed male and female.

Wild animals

This study did not involve wild animals.

Field-collected samples

The study did not involve samples collected from the field.

Ethics oversight

Mouse studies were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Institutional Animal Care and Use Committee at the Washington University School of Medicine (NIH/PHS Assurance ID: D16-00245) and approved by the Institutional Animal Care and Use Committee at the UNC Chapel Hill School of Medicine (NIH/PHS Assurance ID: D16-00256). Virus inoculations were performed under anesthesia that was induced and maintained with ketamine hydrochloride and xylazine, and all efforts were made to minimize animal suffering. The NHP research studies adhered to principles stated in the eighth edition of the Guide for the Care and Use of Laboratory Animals. The facility where this research was conducted (Bioqual Inc., Rockville, MD) is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC) and approved by the Office of Laboratory Animal Welfare (NIH/PHS Assurance ID: D16-00052). NHP studies were conducted in compliance with all relevant local, state, and federal regulations and were approved by the Animal Care and Use Committee (IACUC) at Bioqual.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	We studied 4 subjects with previous laboratory-confirmed symptomatic SARS-CoV-2 infection that was acquired in China, and one healthy control subject: Subject 1: Male, 35 years old Subject 2: Female, 52 years old Subject 3: Male, 56 years old Subject 4: Female, 56 years old Healthy control subject: Subject 5, Male, 58 years old Two subjects from which mAbs were isolated (the 56-year-old male and a 56-year-old female) are a married couple and residents of Wuhan, China, who traveled to Toronto, Canada and were diagnosed with SARS-CoV-2 infection by RT-PCR as described previously (PMID: 32511414). Male subject developed symptoms suggestive of COVID-19 and female subject was asymptomatic when RT-PCR tested. At the time of PBMCs collection, male subject was free of symptoms suggestive of COVID-19 for at least 14 days and both subjects had negative nasopharyngeal swab RT-PCR tests. These samples were transferred to Vanderbilt University Medical Center in Nashville, TN, USA on March 14, 2020.
Recruitment	Study participants were recruited at the hospital in Toronto, and PBMCs were obtained by leukapheresis on March 10, 2020, which is 50 days after symptom onset for the male subject and 18 days after negative RT-PCR test for the female subject. These two subjects were selected on the basis of high SARS-CoV-2-specific B cell frequency in these samples with the aim to facilitate identification of potent monoclonal antibodies, as described previously (PMID: 32511414). Samples were obtained after written informed consent. There was no potential self-selection bias in recruiting patients.
Ethics oversight	Ethics oversight Studies to obtain specimens after written informed consent had been approved by the Institutional Review Board of Vanderbilt University Medical Center, the Institutional Review Board of the University of Washington, and the Research Ethics Board of the University of Toronto.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike

https://doi.org/10.1038/s41586-020-2571-7

Received: 15 June 2020

Accepted: 15 July 2020

Lihong Liu[1,12], Pengfei Wang[1,12], Manoj S. Nair[1,12], Jian Yu[1,12], Micah Rapp[2,12], Qian Wang[3,12], Yang Luo[1], Jasper F.-W. Chan[4,5], Vincent Sahi[1], Amir Figueroa[6], Xinzheng V. Guo[7], Gabriele Cerutti[2], Jude Bimela[2], Jason Gorman[8], Tongqing Zhou[8], Zhiwei Chen[4,5,9], Kwok-Yung Yuen[4,5], Peter D. Kwong[8,10], Joseph G. Sodroski[3], Michael T. Yin[11], Zizhang Sheng[1,2], Yaoxing Huang[1]✉, Lawrence Shapiro[1,2,10]✉ & David D. Ho[1]✉

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic continues, with devasting consequences for human lives and the global economy[1,2]. The discovery and development of virus-neutralizing monoclonal antibodies could be one approach to treat or prevent infection by this coronavirus. Here we report the isolation of sixty-one SARS-CoV-2-neutralizing monoclonal antibodies from five patients infected with SARS-CoV-2 and admitted to hospital with severe coronavirus disease 2019 (COVID-19). Among these are nineteen antibodies that potently neutralized authentic SARS-CoV-2 in vitro, nine of which exhibited very high potency, with 50% virus-inhibitory concentrations of 0.7 to 9 ng ml^{-1}. Epitope mapping showed that this collection of nineteen antibodies was about equally divided between those directed against the receptor-binding domain (RBD) and those directed against the N-terminal domain (NTD), indicating that both of these regions at the top of the viral spike are immunogenic. In addition, two other powerful neutralizing antibodies recognized quaternary epitopes that overlap with the domains at the top of the spike. Cryo-electron microscopy reconstructions of one antibody that targets the RBD, a second that targets the NTD, and a third that bridges two separate RBDs showed that the antibodies recognize the closed, 'all RBD-down' conformation of the spike. Several of these monoclonal antibodies are promising candidates for clinical development as potential therapeutic and/or prophylactic agents against SARS-CoV-2.

The novel coronavirus SARS-CoV-2[1,2] has caused more than 14 million confirmed infections globally, and has caused more than 600,000 deaths. This pandemic has also put much of the world on pause, with unprecedented disruption of lives and unparalleled damage to the economy. A return to some semblance of normality will depend on the ability of science to deliver an effective solution, and the scientific community has responded admirably. Drug development is well underway, and vaccine candidates have entered clinical trials. Another promising approach is the isolation of SARS-CoV-2-neutralizing monoclonal antibodies (mAbs) that could be used as therapeutic or prophylactic agents. The primary target for such antibodies is the viral spike, a trimeric protein[3,4] that is responsible for binding of the virus to the ACE2 receptor on the host cell[1,3,5,6]. The spike protein is comprised of two subunits. The S1 subunit has two major structural elements, RBD and NTD; the S2 subunit mediates virus–cell membrane fusion after the RBD has engaged ACE2. Reports of the discovery of neutralizing mAbs that target the RBD have been published recently[7–11]. We now describe our efforts in isolating and characterizing a collection of mAbs that not only target multiple epitopes on the viral spike but also show very high potency in neutralizing SARS-CoV-2.

Patient selection

Forty patients with PCR-confirmed SARS-CoV-2 infection were enrolled in a cohort study on virus-neutralizing antibodies. Plasma samples from all participants were first tested for neutralizing activity against SARS-CoV-2 pseudovirus (Wuhan-Hu-1 spike pseudotyped with vesicular stomatitis virus). Neutralizing titres varied widely, with half-maximal inhibitory concentrations (IC$_{50}$s) ranging from a reciprocal plasma dilution of less than 100 to roughly 13,000 (Fig. 1a). We selected five patients for isolation of mAbs because their plasma virus-neutralizing titres were

---

[1]Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA. [2]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [3]Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [4]State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, Hong Kong, China. [5]Centre for Virology, Vaccinology and Therapeutics, Health@InnoHK, The University of Hong Kong, Hong Kong Special Administrative Region, Hong Kong, China. [6]Department of Microbiology & Immunology Flow Cytometry Core, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA. [7]Human Immune Monitoring Core, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA. [8]Vaccine Research Center, National Institutes of Health, Bethesda, MD, USA. [9]AIDS Institute, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, Hong Kong, China. [10]Department of Biochemistry, Columbia University, New York, NY, USA. [11]Division of Infectious Diseases, Department of Internal Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA. [12]These authors contributed equally: Lihong Liu, Pengfei Wang, Manoj S. Nair, Jian Yu, Micah Rapp, Qian Wang. ✉e-mail: yh3253@cumc.columbia.edu; lss8@columbia.edu; dh2994@cumc.columbia.edu

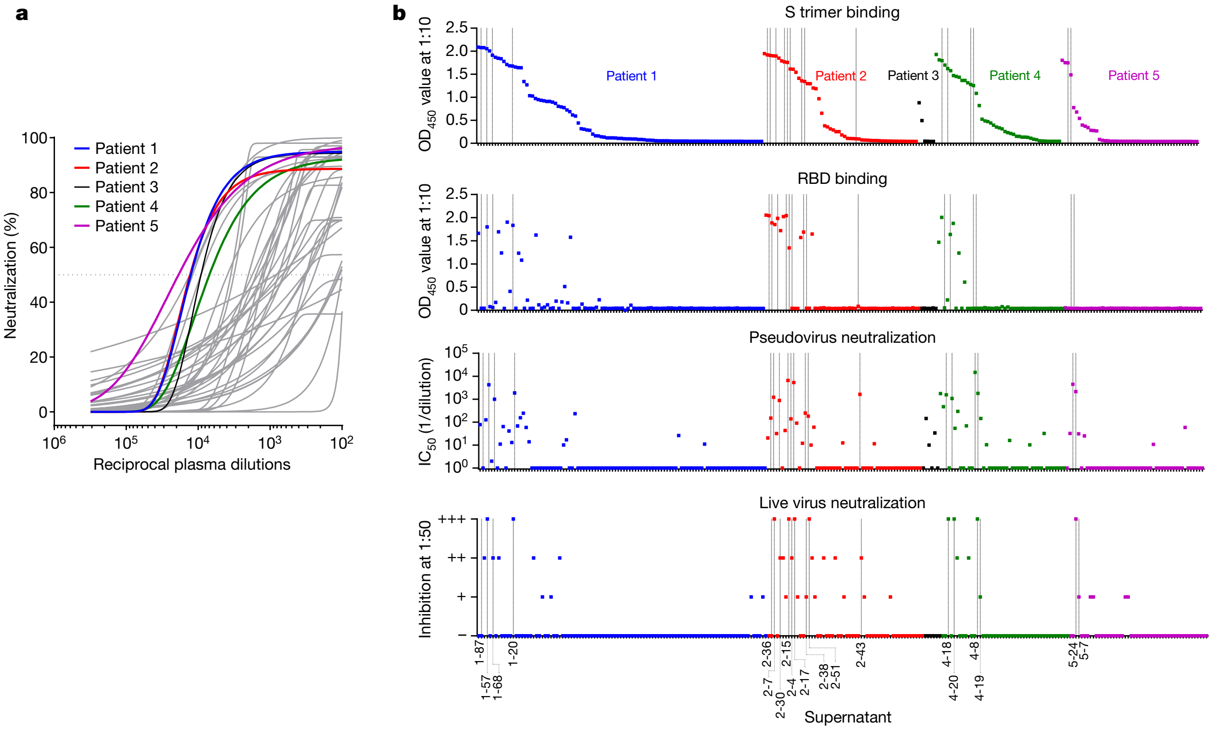


Fig. 1 | Isolation of SARS-CoV-2 mAbs from infected patients with severe disease. a, Plasma neutralization profile of 40 patients against SARS-CoV-2 pseudovirus (highlighted are five top neutralizers chosen for further study). **b**, All 252 transfection supernatants were screened for binding to the S trimer and RBD, as well as for neutralization against SARS-CoV-2 pseudovirus and live virus. For pseudovirus neutralization, the 50% inhibitory dilutions (IC_{50}) of each supernatant are plotted. For live virus, semiquantitative representation of the inhibition at a dilution of 1:50, with neutralization levels ranging from (−) for none to (+++) for complete neutralization, is plotted. Potent antibodies later identified are marked by vertical lines and labelled at the bottom. The antibodies from each patient are coloured as in **a**.

among the highest. The clinical characteristics of these five patients are summarized in Extended Data Table 1. All were severely ill with acute respiratory distress syndrome requiring mechanical ventilation.

Isolation and construction of mAbs

Peripheral blood mononuclear cells from each patient were processed as shown in Extended Data Fig. 1a, starting with cell sorting by flow cytometry. The sorting strategy focused on live memory B lymphocytes that were $CD3^-$, $CD19^+$, and $CD27^+$ (Extended Data Fig. 1b). The final step focused on those cells that bound the SARS-CoV-2 spike trimer (S trimer)[4]. A total of 602, 325, 14, 147, and 145 such B cells from patients 1, 2, 3, 4, and 5, respectively, were labelled with unique hashtags (Extended Data Fig. 1c). The cells were then placed into the 10X Chromium (10X Genomics) for single-cell 5′-mRNA and V(D)J sequencing to obtain paired heavy (H) and light (L) chain sequences. A careful bioinformatic analysis was carried out on 1,145 paired sequences to downselect 'high-confidence' antigen-specific mAbs. We recovered 331 mAb sequences, representing 252 individual clones. Only six mAbs were from patient 3, whereas 44 to 100 mAbs were identified from each of the other patients (Extended Data Fig. 2a). The VH and VL sequences of 252 antibodies (one per clone) were codon-optimized and synthesized, and each VH and VL gene was then cloned into an expression plasmid with corresponding constant regions of H chain and L chain of human IgG1. Monoclonal antibodies were then expressed by co-transfection of paired full-length H chain and L chain genes into Expi293 cells.

Monoclonal antibody screening

All 252 transfection supernatants were screened for binding to the S trimer and RBD by enzyme-linked immunosorbent assays (ELISAs), as well as for their ability to neutralize SARS-CoV-2 pseudovirus and live virus (Fig. 1b, Extended Data Fig. 2). A substantial percentage of the mAbs in the supernatants bound S trimer, and a subset of those bound RBD. Specifically, 121 supernatants were scored as positive for S trimer binding, yielding an overall hit rate of 48%. Of these, 38 were positive for RBD binding while the remaining 83 were negative. None of the 13 trimer-specific mAbs from patient 5 recognized RBD. In the pseudovirus neutralization screen, 61 supernatants were scored as positive, indicating that half of the trimer-specific mAbs were virus-neutralizing. In the screen for neutralization against SARS-CoV-2 (strain USA-WA1/2020), 41 supernatants were scored as positive. Overall, this screening strategy was quite effective in identifying neutralizing mAbs (vertical lines and labelled antibodies at the bottom of Fig. 1b) that were later identified as potent.

Sequence analysis of S trimer-specific mAbs

Of the 121 mAbs that bound the S trimer, 88% were IgG isotype, with IgG1 being predominant (Extended Data Fig. 3a). Comparison to the IgG repertoires of three healthy human donors[12] revealed a statistically significant over-representation of IGHV3-30, IGKV3-20, and IGHJ6 genes for this collection of SARS-CoV-2 mAbs (Extended Data Figs. 3b, c). In addition, the average CDRH3 length was also longer (Extended Data Fig. 3d). Notably, the average percentages of somatic hypermutation in VH and VL were 2.1 and 2.5, respectively, which were significantly lower than those found in healthy individuals (Extended Data Fig. 3e) and remarkably close to those of germline sequences.

Antigen binding and virus neutralization

Since the screening for pseudovirus neutralization was performed quantitatively with serial dilutions of the transfection supernatants,

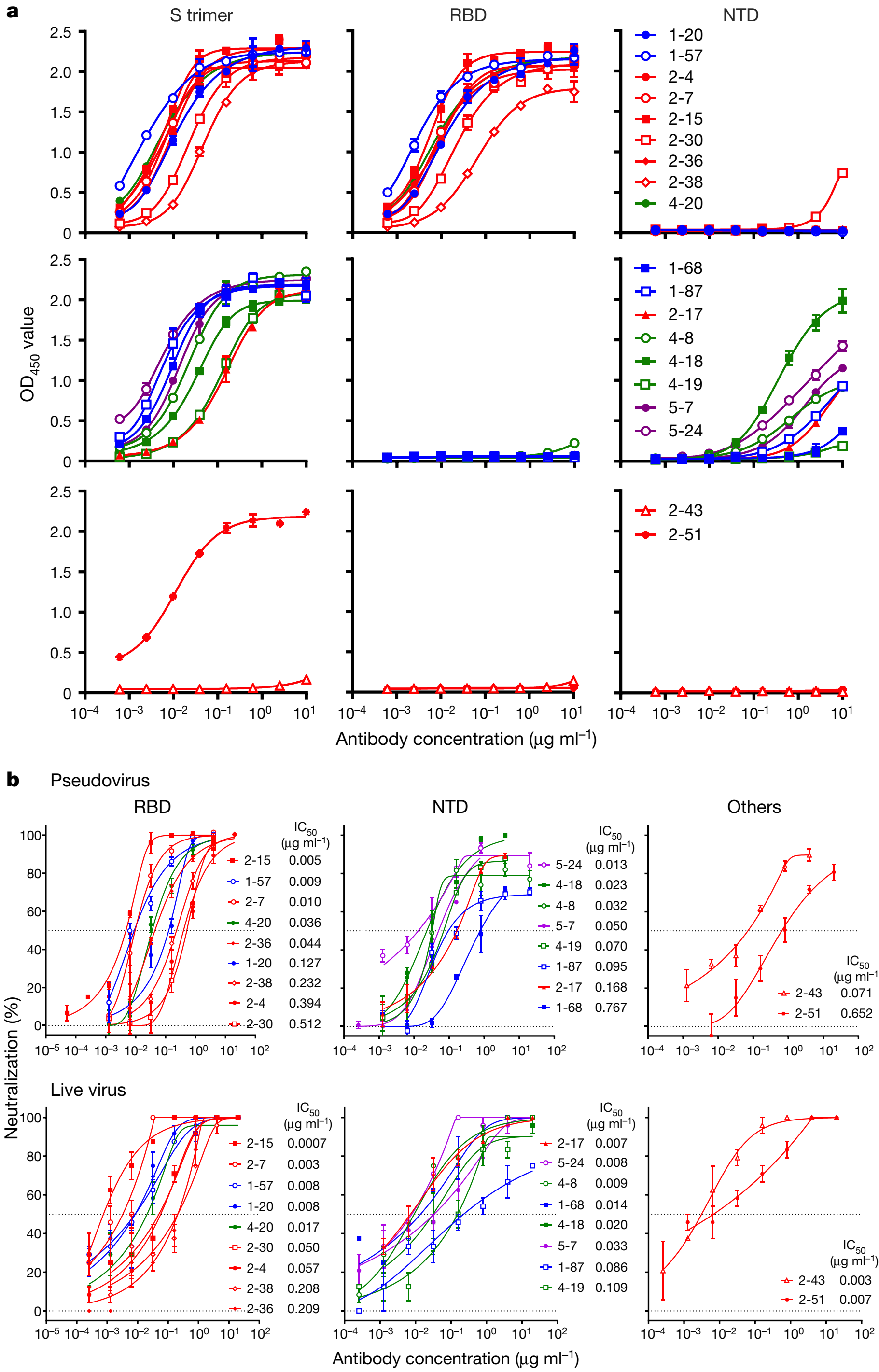
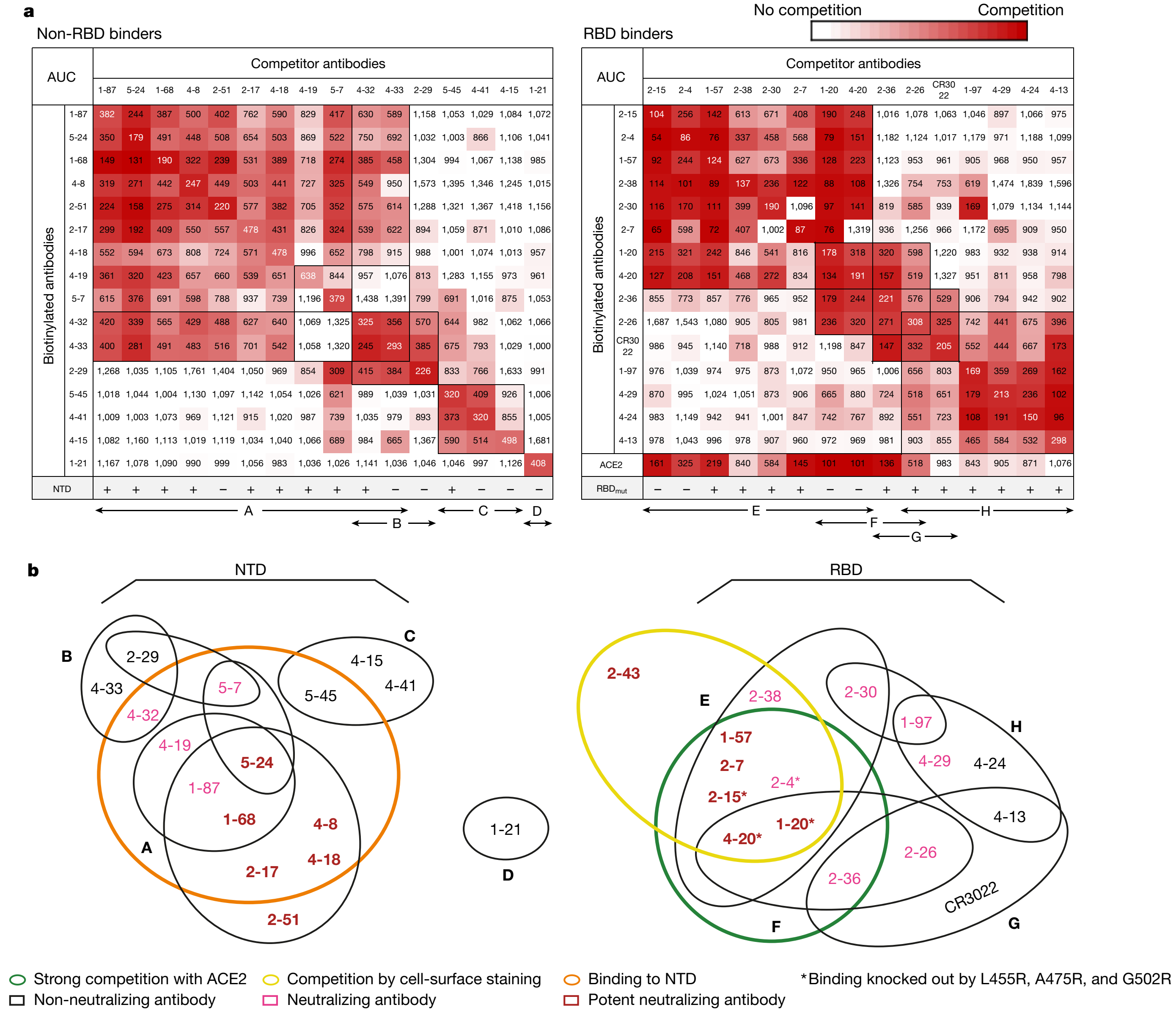


Fig. 2 | Characterization of potent neutralizing mAbs against SARS-CoV-2. **a**, Binding profiles of 19 purified potent neutralizing mAbs against the S trimer (left), RBD (middle), and NTD (right) of SARS-CoV-2. Note that mAb 2-30 bound multiple proteins at high concentrations. **b**, Neutralization profiles of the pseudovirus (top) and live virus (bottom) for the 19 purified mAbs. Epitope classifications are listed above plots. A single replicate of the binding experiment and triplicates of neutralization are presented as mean ± s.e.m.

we plotted in Extended Data Fig. 2b the best-fit neutralization curves for 130 samples that were positive in at least one of the screens shown in Fig. 1b. Most were non-neutralizing or weakly neutralizing, but 18 showed better potency. One additional supernatant was initially missed in the pseudovirus screen (patient 1 in Extended Data Fig. 2b) but was later found to be a potent neutralizing mAb. Together, these 19 mAbs were purified from transfection supernatants and further characterized for their binding and neutralization properties. As shown in Fig. 2a, quantitative ELISA showed that all but one (2-43) of the mAbs bound the S trimer. Nine of the antibodies clearly bound RBD, with little or no binding to NTD. Eight antibodies bound NTD to varying degrees, with no binding to RBD. Two mAbs bound neither RBD nor NTD, and were therefore categorized as ‘other’.

The pseudovirus neutralization profiles for these purified 19 mAbs are shown in Fig. 2b (top). The RBD-directed antibodies neutralized the pseudovirus with IC$_{50}$ values of 0.005 to 0.512 µg ml^{-1}; the NTD-directed



Non-RBD binders

AUC	1-87	5-24	1-68	4-8	2-51	2-17	4-18	4-19	5-7	4-32	4-33	2-29	5-45	4-41	4-15	1-21
1-87	382	244	387	500	402	762	590	829	417	630	589	1,158	1,053	1,029	1,084	1,072
5-24	350	179	491	448	508	654	503	869	522	750	692	1,032	1,003	866	1,106	1,041
1-68	149	131	190	322	239	531	389	718	274	385	458	1,304	994	1,067	1,138	985
4-8	319	271	442	247	449	503	441	727	325	549	950	1,573	1,395	1,346	1,245	1,015
2-51	224	158	275	314	220	577	382	705	352	575	614	1,288	1,321	1,367	1,418	1,156
2-17	299	192	409	550	557	478	431	826	324	539	622	894	1,059	871	1,010	1,086
4-18	552	594	673	808	724	571	478	996	652	798	915	988	1,001	1,074	1,013	957
4-19	361	320	423	657	660	539	651	638	844	957	1,076	813	1,283	1,155	973	961
5-7	615	376	691	598	788	937	739	1,196	379	1,438	1,391	799	691	1,016	875	1,053
4-32	420	339	565	429	488	627	640	1,069	1,325	325	356	570	644	982	1,062	1,066
4-33	400	281	491	483	516	701	542	1,058	1,320	245	293	385	675	793	1,029	1,000
2-29	1,268	1,035	1,105	1,761	1,404	1,050	969	854	309	415	384	226	833	766	1,633	991
5-45	1,018	1,044	1,004	1,130	1,097	1,142	1,054	1,026	621	989	1,039	1,031	320	409	926	1,006
4-41	1,009	1,003	1,073	969	1,121	915	1,020	987	739	1,035	979	893	373	320	855	1,005
4-15	1,082	1,160	1,113	1,019	1,119	1,034	1,040	1,066	689	984	665	1,367	590	514	498	1,681
1-21	1,167	1,078	1,090	990	999	1,056	983	1,036	1,026	1,141	1,036	1,046	1,046	997	1,126	408
NTD	+	+	+	+	−	+	+	+	+	+	−	−	+	−	−	−

RBD binders

AUC	2-15	2-4	1-57	2-38	2-30	2-7	1-20	4-20	2-36	2-26	CR3022	1-97	4-29	4-24	4-13
2-15	104	256	142	613	671	408	190	248	1,016	1,078	1,063	1,046	897	1,066	975
2-4	54	86	76	337	458	568	79	151	1,182	1,124	1,017	1,179	971	1,188	1,099
1-57	92	244	124	627	673	336	128	223	1,123	953	961	905	968	950	957
2-38	114	101	89	137	236	122	88	108	1,326	754	753	619	1,474	1,839	1,596
2-30	116	170	111	399	190	1,096	97	141	819	585	939	169	1,079	1,134	1,144
2-7	65	598	72	407	1,002	87	76	1,319	936	1,256	966	1,172	695	909	950
1-20	215	321	242	846	541	816	178	318	320	598	1,220	983	932	938	914
4-20	127	208	151	468	272	834	134	191	157	519	1,327	951	811	958	798
2-36	855	773	857	776	965	952	179	244	221	576	529	906	794	942	902
2-26	1,687	1,543	1,080	905	805	981	236	320	271	308	325	742	441	675	396
CR3022	986	945	1,140	718	988	912	1,198	847	147	332	205	552	444	667	173
1-97	976	1,039	974	975	873	1,072	950	965	1,006	656	803	169	359	269	162
4-29	870	995	1,024	1,051	873	906	665	880	724	518	651	179	213	236	102
4-24	983	1,149	942	941	1,001	847	742	767	892	551	723	108	191	150	96
4-13	978	1,043	996	978	907	960	972	969	981	903	855	465	584	532	298
ACE2	161	325	219	840	584	145	101	101	136	518	983	843	905	871	1,076
RBD_{mut}	−	−	+	+	+	+	−	−	+	+	+	+	+	+	+

Fig. 3 | Epitope mapping of select neutralizing and non-neutralizing mAbs. **a**, Competition results of non-RBD binders (left) and RBD binders (right) in blocking binding of ACE2 or biotinylated mAb to the S trimer. In addition, the ability of each mAb to bind NTD and RBD_{mut} is shown. The numbers in each box show the area under each competition curve (AUC) as tested by ELISA. Plus and minus signs indicate binding and no binding, respectively, of the mAb to the protein. The letters A to H at the bottom denote clusters of antibody epitopes defined by the competition experiments. **b**, Venn diagram interpretation of results from **a** and Extended Data Fig. 6b.

antibodies were slightly less potent, with IC_{50} values ranging from 0.013 to 0.767 μg ml^{-1}. A common feature of the NTD mAbs was the plateauing of virus neutralization at levels short of 100%. Two antibodies, categorized as 'other', neutralized with IC_{50} values of 0.071 and 0.652 μg ml^{-1}. Antibody neutralization of the authentic or live SARS-CoV-2 (strain USA-WA1/2020) was carried out using Vero cells inoculated with a multiplicity of infection of 0.1. As shown in the bottom portion of Fig. 2b, the RBD-directed antibodies again neutralized the virus but with IC_{50} values of 0.0007 to 0.209 μg ml^{-1}; the NTD-directed antibodies showed similar potency, with IC_{50} values ranging from 0.007 to 0.109 μg ml^{-1}. Here, the plateauing effect seen in the pseudovirus neutralization assay was less apparent. Antibodies 2-43 and 2-51 neutralized the live virus with IC_{50} values of 0.003 and 0.007 μg ml^{-1}, respectively. Overall, nine mAbs exhibited high potency in neutralizing authentic SARS-CoV-2 in vitro with IC_{50} values of 0.009 μg ml^{-1} or less, including four against RBD (2-15, 2-7, 1-57, and 1-20), three against NTD (2-17, 5-24, and 4-8), and two against undetermined regions on the S trimer (2-43 and 2-51). Patient 2 alone contributed five of the top nine SARS-CoV-2 neutralizing mAbs. A correlation of the results of the two virus-neutralizing assays is shown in Extended Data Fig. 4.

Epitope mapping

All 19 potent neutralizing mAbs (Fig. 2) were further studied in antibody competition experiments to gain insight into their epitopes. We also chose 12 mAbs that bound the S trimer strongly during the initial supernatant screen, including 5 that bound RBD (1-97, 2-26, 4-13, 4-24, and 4-29) and 7 that did not bind RBD (1-21, 2-29, 4-15, 4-32, 4-33, 4-41, and 5-45). Four of these mAbs were weak in neutralizing SARS-CoV-2 pseudovirus, and the remaining eight were non-neutralizing (Extended Data

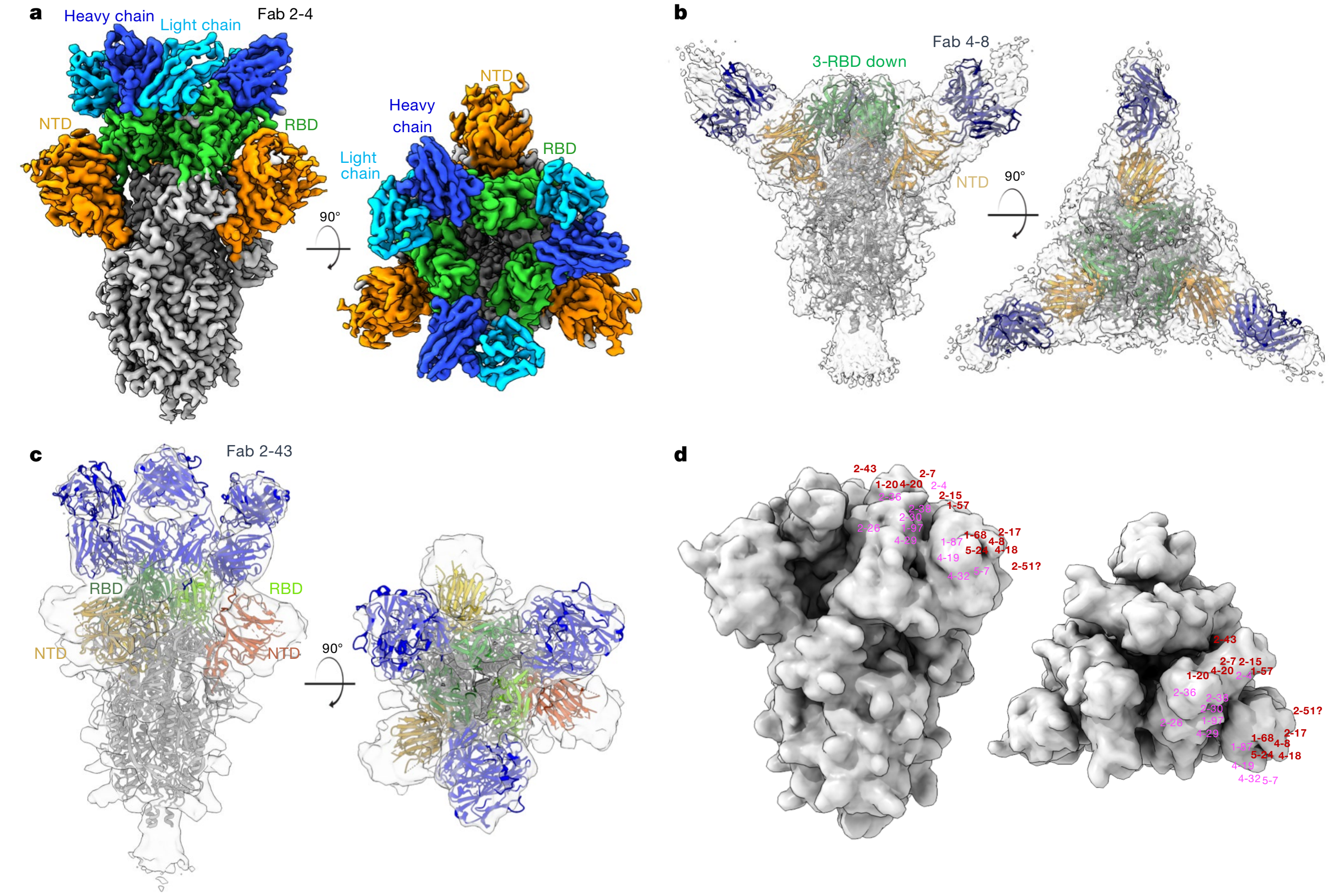


Fig. 4 | Cryo-EM reconstructions of Fab–spike complexes and visualization of neutralizing epitopes on the spike surface. a, Cryo-EM reconstruction of 2-4 Fab in complex with the S trimer at 3.2 Å overall resolution. Density is coloured with RBD in green, NTD in orange, and other regions in grey. **b**, Cryo-EM reconstruction of 4-8 Fab in complex with the S trimer (ribbon diagram, coloured as in **a**) at 3.9 Å overall resolution, with RBDs in the 'all-down' configuration. **c**, Cryo-EM reconstruction of the 2-43 Fab in complex with the S trimer at 5.8 Å resolution reveals a quaternary epitope involving RBD from one subunit and another RBD from the next. **d**, Mapping of the Venn diagrams from Fig. 3b onto the surface of the viral spike.

Fig. 5). We used ELISA to evaluate 16 non-RBD mAbs for competition in binding to the S trimer in a 'checkerboard' experiment. The extent of the antibody competition is reflected by the intensity of the heatmap in Fig. 3a. There is one large cluster (A) of mAbs that competed with one another, which partially overlaps with a small cluster (B). A third cluster (C) does not overlap at all. Note that all but one of the antibodies in cluster A recognized NTD. Antibody 2-51 is clearly directed against the NTD region even though it could not bind NTD. Moreover, one mAb from each of clusters B and C also recognized NTD, thereby indicating that all three clusters are within or near the NTD. One mAb, 1-21, appears to have a unique non-overlapping epitope (epitope region D).

We carried out a similar 'checkerboard' competition experiment by ELISA for 14 of our RBD-directed mAbs plus CR3022[13,14]. Here, the heatmap shows four epitope clusters that are serially overlapping (Fig. 3a). There is one large cluster (E) that contains mAbs that can largely block ACE2 binding. Furthermore, four antibodies in this cluster lost the ability to bind to a mutant RBD (L455R, A475R, G502R) that could no longer bind ACE2 (unpublished data). Together, these results suggest that most of the mAbs in cluster E are directed against the ACE2-binding interface of RBD. The next cluster (F) connects to both cluster E and cluster G, the location of which is defined by its member CR3022[15]. Last, cluster G overlaps another cluster (H), which includes 1-97, which strongly inhibited the binding of 2-30 to the S trimer. This finding suggests that cluster H may be proximal to one edge of cluster E.

One potent neutralizing mAb, 2-43, did not bind the S trimer in ELISA (Fig. 2a) and thus could not be tested in the above competition experiments. However, 2-43 did strongly bind the S trimer when expressed on the cell surface, as determined by flow cytometry (Extended Data Fig. 6a), and this binding was competed out by itself but not by RBD, NTD, ACE2, or the soluble S trimer[4] (Extended Data Fig. 6b). NTD-directed mAbs had only a modest effect on its binding to cell-surface S trimer, but numerous RBD-directed mAbs in cluster E potently blocked the binding of 2-43, demonstrating that this antibody is likely to target a quaternary epitope on the top of RBD.

These mapping results could be represented by two sets of Venn diagrams shown in Fig. 3b. In the non-RBD region, the potent neutralizing mAbs reside exclusively in cluster A and bind a patch on the NTD. Weaker neutralizing mAbs recognize a region at the interface between clusters A and B. In the RBD region, the most potent neutralizing mAbs also group together within one cluster (E). Given that all block ACE2 binding, it is likely they recognize the top of RBD and neutralize the virus by competitive inhibition of receptor binding. Cluster G contains CR3022, a mAb known to be directed against an epitope on a cryptic site on the side of RBD when it is in the 'up' position[15]. Cluster F is therefore likely situated between the top and this 'cryptic' site. The Venn diagram also suggests that cluster H may occupy a different side surface of RBD, perhaps in the region recognized by S309, a mAb isolated from a patient with SARS-CoV-1[8].

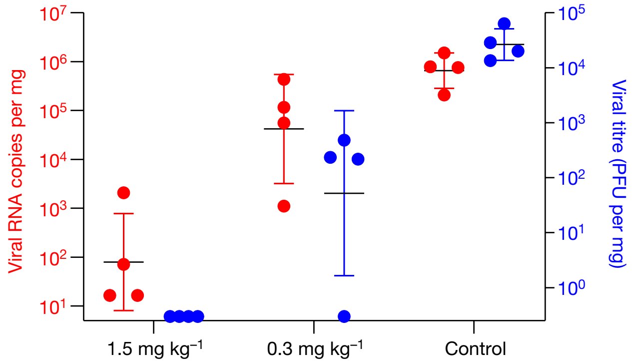


Fig. 5 | Efficacy of mAb 2-15 in protecting against SARS-CoV-2 infection in lung tissues of hamsters. One day before intranasal challenge with SARS-CoV-2, each group of hamsters was given a single intraperitoneal dose of 1.5 mg kg^{-1} of mAb 2-15 ($n = 4$), 0.3 mg kg^{-1} of mAb 2-15 ($n = 4$), or saline as control ($n = 4$). The viral loads in the lung tissues on day 4 after viral challenge were determined by quantitative PCR with reverse transcription (qRT–PCR; red), as well as by an assay to quantify PFUs of infectious SARS-CoV-2 (blue). All data points are shown, along with the mean ± s.d. The differences between the 1.5 mg kg^{-1} group and the control group are statistically significant at $P < 0.05$.

Cryo-electron microscopy

We produced cryo-electron microscopy (cryo-EM) reconstructions of antigen-binding fragments (Fabs) from three mAbs in complex with the S trimer[4]. First, single-particle analysis of the complex with the Fab of mAb 2-4 (RBD-directed) yielded maps of high quality (Fig. 4a; Extended Data Table 2; Extended Data Fig. 7a–d), with the most abundant particle class representing a 3-Fab-per-trimer complex, refined to an overall resolution of 3.2 Å. While density for the constant portion of the Fabs was visible, it was blurred as a result of molecular motion, and thus only the variable domains were included in the molecular model. Fab 2-4 bound the spike protein near the apex, with all RBDs in the 'down' orientation, and the structure of the antibody-bound spike protein was highly similar to previously published unliganded spike structures in the 'all-down' conformation[3,4]. Detailed interactions between mAb 2-4 and RBD are shown in Extended Data Fig. 7e–i. Overall, the structure of the 2-4 Fab–spike complex shows that neutralization of SARS-CoV-2 by this mAb is likely to result from locking the RBD in the down conformation while also occluding access to ACE2.

We also produced 3D cryo-EM reconstructions of 4-8 Fab (NTD-directed) in complex with the S trimer (Extended Data Table 2, Extended Data Fig. 8a–f). Two main particle classes were observed—one for a 3-Fab-bound complex with all RBDs 'down' at 3.9 Å resolution (Fig. 4b), and another a 3-Fab-bound complex with one RBD 'up' at 4.0 Å resolution (Extended Data Fig. 8g). However, molecular motion prevented visualization of the interaction at high resolution. Nevertheless, the density in the 4-8 map reveals the overall positions of the antibody chains that target the NTD. It is unclear how binding to the tip of the NTD results in neutralization of SARS-CoV-2.

Third, a 5.8 Å resolution reconstruction of 2-43 Fab in complex with the S trimer (Extended Data Table 2, Extended Data Fig. 8h–k) revealed three bound Fabs, each targeting a quaternary epitope on the top of the spike that included elements of the RBDs from two adjacent S1 protomers (Fig. 4c), consistent with the epitope mapping results (Fig. 3b, Extended Data Fig. 6b), including the lack of binding to isolated RBD (Fig. 2a). Given these findings, the inability of 2-43 to bind the S trimer in ELISA studies is likely to be an artefact of the assay format, as this mAb did bind the spike expressed on the cell surface and in the cryo-EM study.

Armed with these three cryo-EM reconstructions, we used the Venn diagrams from Fig. 3b to map the epitopes of many of our SARS-CoV-2-neutralizing mAbs onto the surface of the spike (Fig. 4d). This is obviously a rough approximation because antibody footprints are much larger than the area occupied by the mAb number. However, the spatial relationship of the antibody epitopes should be reasonably represented by Fig. 4d.

mAb 2-15 protects hamsters against SARS-CoV-2

To assess the in vivo potency of mAb 2-15, we performed a protection experiment in a golden Syrian hamster model of SARS-CoV-2 infection. The hamsters were first given an intraperitoneal injection of the antibody at a dose of 1.5 mg kg^{-1} or 0.3 mg kg^{-1}, or PBS alone. Intranasal inoculations of 10^5 plaque-forming units (PFU) of the HKU-001a strain of SARS-CoV-2 were carried out 24 h later. Four days after virus challenge, lung tissues were removed to quantify the viral load. As shown in Fig. 5, both viral RNA copy numbers and infectious virus titres were reduced by 4 logs or more in hamsters given 1.5 mg kg^{-1} of mAb 2-15. The protection at 0.3 mg kg^{-1} was borderline, as we had estimated. This pilot animal study demonstrates that the potency of mAb 2-15 in vitro is reflected in vivo, with complete elimination of infectious SARS-CoV-2 at a relatively modest antibody dose.

Discussion

We have identified a collection of SARS-CoV-2-neutralizing mAbs that are not only potent but also diverse. Nine of these antibodies can neutralize the authentic virus in vitro at concentrations of 9 ng ml^{-1} or less (Fig. 2b), including four directed against the RBD, three directed against the NTD, and two directed against nearby quaternary epitopes. Unexpectedly, many of the these mAbs have V(D)J sequences close to germline sequences, without extensive somatic hypermutations (Extended Data Fig. 3e), a finding that bodes well for vaccine development. Our most potent RBD-specific mAbs (for example, 2-15, 2-7, 1-57, and 1-20) compare favourably with such antibodies recently reported[7,8,10,16–20], including those with high potency[9,11,21,22]. The in vitro potency of 2-15 is well reflected in vivo in the hamster protection experiment (Fig. 5). It appears from the epitope-mapping studies that mAbs directed against the top of the RBD compete strongly with ACE2 binding and potently neutralize the virus, whereas those directed against the side surfaces of the RBD do not compete with ACE2 and neutralize less potently (Figs. 3b, 4d). Our collection of non-RBD neutralizing mAbs is unprecedented, to our knowledge, in that such antibodies have been reported only sporadically and only with substantially lower potencies[22–24]. The most potent of these mAbs are directed against (for example, 2-17, 5-24, and 4-8) or overlapping with (2-51) a patch on the NTD (Figs. 3b, 4d). It is unclear how NTD-directed mAbs block SARS-CoV-2 infection and why their neutralization profiles are different from those of RBD-directed antibodies (Fig. 2b). Nevertheless, vaccine strategies that do not include the NTD will be unable to induce an important class of virus-neutralizing antibodies.

The isolation of two mAbs (2-43 and 2-51) directed against epitopes that do not map to the RBD or NTD is also unprecedented, to our knowledge. Cryo-EM of 2-43 Fab bound to the S trimer has confirmed its epitope as quaternary in nature, crossing from the top of one RBD to the top of another RBD (Fig. 4c). It will be equally informative to understand the epitope of 2-51. We have also shown cryo-EM evidence for a neutralizing mAb (4-8) bound to the NTD of the viral spike (Fig. 4b), as well as another high-resolution structure of an mAb (2-4) bound to the RBD (Fig. 4a).

The potency and diversity of our SARS-CoV-2-neutralizing mAbs are probably attributable to patient selection. Infected individuals with severe disease develop a more robust virus-neutralizing antibody response[25]. If patient 2 had not been included, five of the top neutralizing mAbs would have been lost. The diversity of our antibodies is also attributable, in part, to the choice of using the S trimer to sort from memory B cells, while most groups have used the RBD[7,9–11,16–19,21]. The characterization of this diverse collection of mAbs has allowed us to observe that all potent SARS-CoV-2-neutralizing antibodies described to date are directed against the top of the viral spike. RBD and NTD are, undoubtedly, quite immunogenic. Neutralizing antibodies to the stem region of the S trimer remain to be discovered. In conclusion, we believe that several of our monoclonal antibodies with strong virus-neutralizing activity are promising candidates for development as modalities to treat or prevent SARS-CoV-2 infection.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2571-7.

1. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
3. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e286 (2020).
4. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
5. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e278 (2020).
6. Wang, Q. et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **181**, 894–904.e899 (2020).
7. Ju, B. et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* https://doi.org/10.1038/s41586-020-2380-z (2020).
8. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
9. Cao, Y. et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84.e16 (2020).
10. Wu, Y. et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).
11. Hansen, J. et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* https://doi.org/10.1126/science.abd0827 (2020).
12. Sheng, Z. et al. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front. Immunol*. **8**, 537 (2017).
13. ter Meulen, J. et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med*. **3**, e237 (2006).
14. Tian, X. et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect*. **9**, 382–385 (2020).
15. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
16. Rogers, T. F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* https://doi.org/10.1126/science.abc7520 (2020).
17. Chen, X. et al. Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. *Cell. Mol. Immunol*. **17**, 647–649 (2020).
18. Zeng, X. et al. Isolation of a human monoclonal antibody specific for the receptor binding domain of SARS-CoV-2 using a competitive phage biopanning strategy. *Antib. Ther*. **3**, 95–100 (2020).
19. Liu, X. et al. Neutralizing antibodies isolated by a site-directed screening have potent protection on SARS-CoV-2 infection. Preprint at https://doi.org/10.1101/2020.05.03.074914 (2020).
20. Zost, S. J. et al. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat. Med*. https://doi.org/10.1038/s41591-020-0998-x (2020).
21. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals. *Nature* https://doi.org/10.1038/s41586-020-2456-9 (2020).
22. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* https://doi.org/10.1126/science.abc5902 (2020).
23. Chi, X. et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* https://doi.org/10.1126/science.abc6952 (2020).
24. Wang, C. et al. A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun*. **11**, 2251 (2020).
25. Wang, P. et al. SARS-CoV-2 neutralizing antibody responses are more robust in patients with severe disease. Preprint at https://doi.org/10.1101/2020.06.13.150250 (2020).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment, except where stated.

Expression and purification of SARS-CoV-2 proteins

The mammalian expression vector that encodes the ectodomain of the SARS-CoV-2 S trimer and the vector encoding RBD fused with SD1 at the N terminus and an HRV-3C protease cleavage site followed by a mFc tag and an 8 × His tag at the C terminus were kindly provided by Jason McLellan[4]. SARS-CoV-2 NTD (aa1-290) with an HRV-3C protease cleavage site, a mFc tag, and an 8 × His tag at the C terminus was also cloned into mammalian expression vector pCAGGS. Each expression vector was transiently transfected into Expi293 cells using 1 mg/ml polyethylenimine (Polysciences). Five days after transfection, the S trimer was purified using Strep-Tactin XT Resin (Zymo Research), and the RBD-mFc and NTD-mFc were purified using protein A agarose (ThermoFisher Scientific). In order to obtain RBD-SD1 and NTD, the mFc and 8 × His tags at the C terminus were removed by HRV-3C protease (Millipore-Sigma) and then purified using Ni-NTA resin (Invitrogen) followed by protein A agarose.

Sorting for S trimer-specific B cells and single-cell B cell receptor sequencing

Peripheral blood mononuclear cells from five patients and one healthy donor were stained with LIVE/DEAD Fixable Yellow Dead Cell Stain Kit (Invitrogen) at ambient temperature for 20 min, followed by washing with RPMI-1640 complete medium and incubation with 10 µg/ml S trimer at 4 °C for 45 min. Afterwards, the cells were washed again and incubated with a cocktail of flow cytometry and hashtag antibodies, containing CD3 PE-CF594 (BD Biosciences), CD19 PE-Cy7 (Biolegend), CD20 APC-Cy7 (Biolegend), IgM V450 (BD Biosciences), CD27 PerCP-Cy5.5 (BD Biosciences), anti-His PE (Biolegend), and human Hashtag 3 (Biolegend) at 4 °C for 1 h. Stained cells were then washed, resuspended in RPMI-1640 complete medium and sorted for S trimer-specific memory B cells ($CD3^{-}CD19^{+}CD27^{+}$S trimer^{+} live single lymphocytes). The sorted cells were mixed with mononuclear cells from the same donor, labelled with Hashtag 1, and loaded into the 10X Chromium chip of the 5′ Single Cell Immune Profiling Assay (10X Genomics) at the Columbia University Human Immune Monitoring Core (HIMC; RRID:SCR_016740). The library preparation and quality control were performed according to the manufacturer's protocol and sequenced on a NextSeq 500 sequencer (Illumina).

Identification of S trimer-specific antibody transcripts

For each sample, full-length antibody transcripts were assembled using the VDJ module in Cell Ranger (version 3.1.0, 10X Genomics) with default parameters and the GRCh38 genome as reference. To identify cells from the antigen sort, we first used the count module in Cell Ranger to calculate copies of all hashtags in each cell from the Illumina NGS raw reads. High-confidence antigen-specific cells were identified as follows. In brief, based on the copy numbers of the hashtags observed, a cell must contain more than 100 copies of the antigen sort-specific hashtag to qualify as an antigen-specific cell. Because hashtags can fall off cells and bind to cells from a different population in the sample mixture, each cell usually has both sorted and spiked-in-specific hashtags. To enrich for true antigen-specific cells, the copy number of the specific hashtag has to be at least 1.5× higher than that of the non-specific hashtag. Low-quality cells were identified and removed using the cell-calling algorithm in Cell Ranger. Cells that did not have productive H and L chain pairs were excluded. If a cell contained more than two H or/and L chain transcripts, the transcripts with fewer than three unique molecular identifiers were removed. Cells with identical H and L chain sequences, which may have resulted from mRNA leakage, were merged into one cell. Additional filters were applied to remove low-quality cells and/or transcripts in the antibody gene annotation process.

Antibody transcript annotation and selection criteria

Antigen-specific antibody transcripts were processed using our bioinformatics pipeline SONAR for quality control and annotation[26]. In brief, V(D)J genes were assigned for each transcript using BLAST[27] with customized parameters against a germline gene database obtained from the international ImMunoGeneTics information system (IMGT) database[26,28]. On the basis of BLAST alignments of V and J regions, CDR3 was identified using the conserved second cysteine in the V region and WGXG (H chain) or FGXG (L chain) motifs in the J region (X represents any amino acid). For H chain transcripts, the constant domain 1 (CH1) sequences were used to assign isotype using BLAST with default parameters against a database of human CH1 genes obtained from IMGT. A BLAST E-value threshold of 10^{-6} was used to find significant isotype assignments, and the CH1 allele with the lowest E-value was used. Sequences other than the V(D)J region were removed and transcripts containing incomplete V(D)J or/and frame shift were excluded. We then aligned each of the remaining transcripts to the assigned germline V gene using CLUSTALO[29] and calculated the somatic hypermutation level.

To select representative antibodies for functional characterization, we first clustered all antibodies using USEARCH[30] with the following criteria: identical heavy chain V and J gene assignments, the same length of CDRH3, and CDRH3 identity higher than 0.9. For each cluster, cells with the same light chain V and J gene assignments were grouped into a clone. All clone assignments were manually checked. We then calculated the clonal size for each clone, and one H and L chain pair per clone was chosen for antibody synthesis. For clones with multiple members, the member with the highest somatic hypermutation level was chosen for synthesis. For cells having multiple high quality H or L chains, which may be from doublets, we synthesized all H and L chain combinations.

Analysis of S trimer-specific antibody repertoire

Because 88% of the S trimer-specific antibodies were IgG isotype, we compared the repertoire features to IgG repertoires from three healthy donors[12] (17,243 H chains, 27,575 kappa L chains, 20,889 lambda L chains). The repertoire data from the three healthy donors were combined and annotated using SONAR with the same process as above.

Antibody expression and purification

For each antibody, variable genes were optimized for human cell expression and synthesized by GenScript. VH and VL were inserted separately into plasmids (gWiz or pcDNA3.4) that encode the constant region for H chain and L chain. Monoclonal antibodies were expressed in Expi293 (ThermoFisher, A14527) by co-transfection of H chain and L chain expressing plasmids using polyethylenimine and culture at 37 °C with shaking at 125 rpm and 8% CO_2. On day 3 after transfection, 400 µl supernatant were collected for screening for binding to the S trimer and RBD by ELISA, and for neutralization of SARS-CoV-2 pseudovirus and authentic virus. Supernatants were also collected on day 5 for antibody purification using rProtein A Sepharose (GE, 17-1279-01) affinity chromatography.

Production of pseudoviruses

Recombinant Indiana vesicular stomatitis virus (rVSV) expressing the SARS-CoV-2 spike was generated as previously described[31,32]. HEK293T cells were grown to 80% confluency before transfection with pCMV3-SARS-CoV-2-spike (Sino Biological) using FuGENE 6 (Promega). Cells were cultured overnight at 37 °C with 5% CO_2. The next day, medium was removed and VSV-G pseudotyped ΔG-luciferase (G*ΔG-luciferase, Kerafast) was used to infect the cells in DMEM at a

MOI of 3 for 1 h before the cells were washed three times with 1× DPBS. DMEM supplemented with 2% fetal bovine serum, 100 IU/ml of penicillin and 100 μg/ml of streptomycin were added to the inoculated cells, which were cultured overnight as described above. The supernatant was removed the following day and clarified by centrifugation at 300*g* for 10 min before aliquoting and storing at −80 °C.

Pseudovirus neutralization

Neutralization assays were performed by incubating pseudoviruses with serial dilutions of heat-inactivated plasma together with supernatant or purified antibodies, and scored by the reduction in luciferase gene expression. In brief, Vero E6 cells (ATCC) were seeded in a 96-well plate at a concentration of 2×10^4 cells per well. Pseudoviruses were incubated the next day with serial dilutions of the test samples in duplicate or triplicate for 30 min at 37 °C. The mixture was added to cultured cells and incubated for an additional 24 h. The luminescence was measured using a Britelite plus Reporter Gene Assay System (PerkinElmer). IC_{50} was defined as the dilution at which the relative light units were reduced by 50% compared with the virus control wells (virus + cells) after subtraction of the background in the control groups with cells only. The IC_{50} values were calculated using nonlinear regression in GraphPad Prism 8.0.

Authentic SARS-CoV-2 neutralization

Supernatants containing expressed mAbs were diluted 1:10 and 1:50 in EMEM with 7.5% inactivated fetal calf serum and incubated with authentic SARS-CoV-2 (strain USA-WA1/2020; MOI 0.1) for 1 h at 37 °C. After incubation, the mixture was transferred onto a monolayer of Vero-E6 cells that was cultured overnight. After incubation of the cells with the mixture for 70 h at 37 °C, cytopathic effects (CPEs) caused by the infection were scored for each well from 0 to 4 to indicate the degree of virus inhibition. Semi-quantitative representation of the inhibition for each antibody-containing supernatant at a dilution of 1:50 is shown in the lowest panel of Fig. 1b with neutralization levels ranging from (−) for none to (+++) for complete neutralization.

An end-point dilution assay in a 96-well plate format was performed to measure the neutralization activity of select purified mAbs. In brief, each antibody was serially diluted (fivefold dilutions) starting at 20 μg/ml. Triplicates of each mAb dilution were incubated with SARS-CoV-2 at a MOI of 0.1 in EMEM with 7.5% inactivated fetal calf serum for 1 h at 37 °C. After incubation, the virus–antibody mixture was transferred onto a monolayer of Vero-E6 cells grown overnight. The cells were incubated with the mixture for 70 h. CPEs were visually scored for each well in a blinded fashion by two independent observers. The results were then converted into percentage neutralization at a given mAb concentration, and the averages ± s.e.m. were plotted using a five-parameter dose–response curve in GraphPad Prism 8.0.

Epitope mapping by ELISA

We coated 50 ng per well of S trimer, 50 ng per well of RBD, and 100 ng per well of NTD onto ELISA plates at 4 °C overnight. The ELISA plates were then blocked with 300 μl blocking buffer (1% BSA and 10% bovine calf serum (BCS) (Sigma)) in PBS at 37 °C for 2 h. Afterwards, supernatants from the antibody transfection or purified antibodies were serially diluted using dilution buffer (1% BSA and 20% BCS in PBS), incubated at 37 °C for 1 h. Next, 100 μl of 10,000-fold diluted Peroxidase AffiniPure goat anti-human IgG (H+L) antibody (Jackson ImmunoResearch) was added into each well and incubated for 1 h at 37 °C. The plates were washed between each step with PBST (0.5% Tween-20 in PBS). Finally, the TMB substrate (Sigma) was added and incubated before the reaction was stopped using 1 M sulfuric acid. Absorbance was measured at 450 nm.

For the competition ELISA, purified mAbs were biotin-labelled using One-Step Antibody Biotinylation Kit (Miltenyi Biotec) following the manufacturer's recommendations and purified using 40K MWCO Desalting Column (ThermoFisher Scientific). Serially diluted competitor antibodies (50 μl) were added into S trimer-precoated ELISA plates, followed by 50 μl of biotinylated antibodies at a concentration that achieves an OD_{450} reading of 1.5 in the absence of competitor antibodies. Plates were incubated at 37 °C for 1 h, and 100 μl of 500-fold diluted Avidin-HRP (ThermoFisher Scientific) was added into each well and incubated for another 1 h at 37 °C. The plates were washed with PBST between each of the previous steps. The plates were developed afterwards with TMB and absorbance was read at 450 nm after the reaction was stopped.

For the ACE2 competition ELISA, 100 ng of ACE2 protein (Abcam) was immobilized on the plates at 4 °C overnight. The unbound ACE2 was washed away by PBST and then the plates were blocked. After washing, 100 ng of S trimer in 50 μl dilution buffer was added into each well, followed by addition of another 50 μl of serially diluted competitor antibodies and then incubation at 37 °C for 1 h. The ELISA plates were washed four times with PBST and then 100 μl of 2,000-fold diluted anti-strep-HRP (Millipore Sigma) was added into each well for another 1 h at 37 °C. The plates were then washed and developed with TMB, and absorbance was read at 450 nm after the reaction was stopped.

For all the competition ELISA experiments, the relative binding of biotinylated antibodies or ACE2 to the S trimer in the presence of competitors was normalized by comparing to competitor-free controls. Relative binding curve and the area under curve (AUC) were generated by fitting the nonlinear five-parameter dose–response curve in GraphPad Prism 8.0.

Cell-surface competition binding assay

Expi293 cells were co-transfected with vectors encoding pRRL-cPPT-PGK-GFP (Addgene) and pCMV3-SARS-CoV-2 (2019-nCoV) Spike (Sino Biological) at a ratio of 1:1. Two days after transfection, cells were incubated with a mixture of biotinylated mAb 2-43 (0.25 μg/ml) and serially diluted competitor antibodies at 4 °C for 1 h. Then 100 μl of diluted APC-streptavidin (Biolegend) was added to the cells and incubated at 4 °C for 45 min. Cells were washed three times with FACS buffer before each step. Finally, cells were resuspended and binding of 2-43 to cell-surface S trimer was quantified on an LSRII flow cytometer (BD Biosciences). The mean fluorescence intensity of APC in GFP-positive cells was analysed using FlowJo and the relative binding of 2-43 to the S trimer in the presence of competitors was calculated as the percentage of the mean fluorescence intensity compared to that of the competitor-free controls.

Cryo-EM data collection and processing

SARS-CoV-2 S trimer at a final concentration of 2 mg/ml was incubated with sixfold molar excess per spike monomer of the antibody Fab fragments for 30 min in 10 mM sodium acetate pH 5.5, 150 mM NaCl, and 0.005% *n*-dodecyl-β-D-maltoside (DDM). Sample (2 μl) was incubated on C-flat 1.2/1.3 carbon grids for 30 s and vitrified using a Leica EM GP Plunge Freezer. Data were collected on a Titan Krios electron microscope operating at 300 kV equipped with a Gatan K3 direct detector and energy filter using the Leginon software package[33]. A total electron fluence of 51.3 e/$Å^2$ was fractionated over 40 frames, with a total exposure time of 2 s. A magnification of 81,000× resulted in a pixel size of 1.058 Å, and a defocus range of −0.4 to −3.5 μm was used. All processing was done using cryoSPARC v2.14.2[34]. Raw movies were aligned and dose-weighted using patch motion correction, and the CTF was estimated using patch CTF estimation. A small subset of approximately 200 micrographs were picked using blob picker, followed by 2D classification and manual curation of particle picks, and used to train a Topaz neural network[35]. This network was then used to pick particles from the remaining micrographs, which were extracted with a box size of 384 pixels.

For the 2-4 Fab dataset, 2D classification followed by ab initio modelling and 3D heterogeneous refinement revealed 83,927 particles with three 2-4 Fabs bound, one to each RBD. A reconstruction of these particles using non-uniform refinement with imposed C3 symmetry resulted in a 3.6 Å map, as determined by the gold standard Fourier shell correlation (FSC). Given the relatively low resolution of the RBD–Fab interface, masked local refinement was used to obtain a 3.5 Å map with improved density. A masked local refinement of the remainder of the S trimer resulted in a 3.5 Å reconstruction. These two local refinements were aligned and combined using the vop maximum function in UCSF Chimera[36]. This was repeated for the half maps, which were used, along with the refinement mask from the global non-uniform refinement, to calculate the 3D FSC[37] and obtain an estimated resolution of 3.2 Å. All maps have been submitted to the EMDB with the ID EMD-22156.

For the 4-8 Fab dataset, image preprocessing and particle picking were performed as above. 2D classification, ab initio modelling, and 3D heterogeneous classification revealed 47,555 particles with 3 Fabs bound, one to each NTD and with all 3 RBDs in the down conformation. While this particle stack was refined to 3.9 Å using non-uniform refinement with imposed C3 symmetry, substantial molecular motion prevented the visualization of the Fab epitope at high resolution (EMD-22159). In addition, 105,278 particles were shown to have 3 Fabs bound, but with 1 RBD in the up conformation. These particles were refined to 4.0 Å using non-uniform refinement with C1 symmetry (EMD-22158), and suffered from the same conformational flexibility as the all-RBD-down particles. This flexibility was visualized using 3D variability analysis in cryoSPARC.

For the 2-43 Fab dataset, which was collected at an electron fluence of 51.69 e/Å^2, image preprocessing was performed as above, and particle picking was performed using blob picker. 2D classification, ab initio modelling, and 3D heterogeneous classification revealed 10,068 particles with 3 Fabs bound, which was refined to 5.8 Å resolution (EMD-22157).

Cryo-EM model fitting

An initial homology model of the 2-4 Fab was built using Schrodinger Release 2020-2: BioLuminate[38]. The RBD was initially modelled using the coordinates from PDB ID 6W41. The remainder of the S trimer was modelled using the coordinates from PDB ID 6VSB. These models were docked into the consensus map using Chimera. The model was then fitted interactively using ISOLDE 1.0b5[39] and COOT 0.8.9.2[40], and using real space refinement in Phenix 1.18[41]. In cases where side chains were not visible in the experimental data, they were truncated to alanine. Validation was performed using Molprobity[42] and EMRinger[43]. The model was submitted to the PDB with the ID 6XEY. Figures were prepared using ChimeraX[44].

Hamster protection experiment

In vivo evaluation of mAb 2-15 in an established golden Syrian hamster model of SARS-CoV-2 infection was performed as described previously with slight modifications[45]. Approval was obtained from the University of Hong Kong (HKU) Committee on the Use of Live Animals in Teaching and Research. In brief, 6–8-week-old male and female hamsters were obtained from the Chinese University of Hong Kong Laboratory Animal Service Centre through the HKU Laboratory Animal Unit and kept in biosafety level-2 (BSL-2) housing with access to standard pellet feed and water ad libitum until virus challenge in the BSL-3 animal facility. Each hamster ($n = 4$ per group) was intraperitoneally administered one dose of 1.5 mg/kg of mAb 2-15 in phosphate-buffered saline (PBS), 0.3 mg/kg of mAb 2-15 in PBS, or PBS alone as control. Twenty-four hours later, each hamster was intranasally inoculated with a challenge dose of 100 μl Dulbecco's modified Eagle medium containing 10^5 PFU of SARS-CoV-2 (HKU-001a strain, GenBank accession no: MT230904.1) under intraperitoneal ketamine (200 mg/kg) and xylazine (10 mg/kg) anaesthesia. The hamsters were monitored twice daily for clinical signs of disease and killed on the fourth day after the challenge. Half of each hamster's lung tissue was used for viral load determination by a quantitative SARS-CoV-2 RdRp/Hel RT–PCR assay[46] and an infectious virus titration using a plaque assay described previously[45]. Student's *t*-test was used to determine significant differences among the groups, and $P < 0.05$ was considered statistically significant.

Ethics statement

The acquisition of samples from recovering patients for isolation and identification of potent monoclonal antibodies against COVID-19 (AAAS9517) was approved by the Columbia University Institutional Review Board. Informed consent was obtained from all participants or surrogates.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The 19 neutralizing antibodies have been deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) with accession numbers from MT712278 to MT712315. Coordinates for the antibody 2-4 complex have been deposited in the Protein Data Bank as PDB 6XEY. Cryo-EM maps and data have been deposited in EMDB with deposition codes EMDB-22156 for antibody 2-4, EMDB-22158 and EMDB-22159 for antibody 4-8, and EMDB-22275 for antibody 2-43. These data are used in Fig. 4 and Extended Data Figs. 7, 8.

Code availability

Next-generation sequencing data of antibody repertoires were processed using Cell ranger v3.1.0, SONAR V1, BLAST v2.2.25, CLUSTALO1.2.3, and USEARCH v9.2.64. Cryo-EM data was collected using Leginon 3.4.beta. Cryo-EM data was processed using cryoSPARC v2.14.2, MotionCor2, Topaz v0.2.4, 3DFSC v3.0, UCSF Chimera v1.13.1, ChimeraX v0.93, ISOLDE v1.0b5, Phenix v1.18, and COOT v0.8.9.2.

26. Schramm, C. A. et al. SONAR: a high-throughput pipeline for inferring antibody ontogenies from longitudinal sequencing of B cell transcripts. *Front. Immunol.* **7**, 372 (2016).
27. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
28. Lefranc, M. P. et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* **37**, D1006–D1012 (2009).
29. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116 (2014).
30. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
31. Nie, J. et al. Establishment and validation of a pseudovirus neutralization assay for SARS-CoV-2. *Emerg. Microbes Infect.* **9**, 680–686 (2020).
32. Whitt, M. A. Generation of VSV pseudotypes using recombinant ΔG-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines. *J. Virol. Methods* **169**, 365–374 (2010).
33. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
34. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
35. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
36. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
37. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
38. Zhu, K. et al. Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins* **82**, 1646–1655 (2014).
39. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
41. Adams, P. D. et al. Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11**, 53–55 (2004).
42. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
43. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

44. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. **27**, 14–25 (2018).
45. Chan, J. F. et al. Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa325 (2020).
46. Chan, J. F. et al. Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated *in vitro* and with clinical specimens. *J. Clin. Microbiol*. **58**, e00310-20 (2020).

Acknowledgements We thank N. Wang and J. McLellan for providing reagents for generating the SARS-CoV-2 S trimer and RBD-SD1 with mFc tag; W. Chen for assistance with generating the Venn diagrams; and B. DeKosky and X. Wu for helpful input. This study was supported by funding to D.D.H. from the Jack Ma Foundation, the JPB Foundation, Samuel Yin, Brii Biosciences, Tencent Charity Foundation, Roger Wu, Carol Ludwig, and Peggy and Andrew Cherng. Cryo-EM data collection was performed at the National Center for CryoEM Access and Training and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129539) and by grants from the Simons Foundation (SF349247) and NY State Assembly. Data analysis was performed at the National Resource for Automated Molecular Microscopy (NRAMM), supported by the NIH National Institute of General Medical Sciences (GM103310). Hamster experiments were conducted with support from the Health@InnoHK (Centre for Virology, Vaccinology and Therapeutics), Innovation and Technology Commission, The Government of the Hong Kong Special Administrative Region.

Author contributions D.D.H. conceived the project. L.L., P.W., M.S.N, J.Y., Q.W. and Y.H. performed many of the experiments. M.T.Y. was responsible for recruiting patients, obtaining clinical specimens, and summarizing clinical data. L.L., V.S., A.F. and X.V.G. performed and analysed the B cell sorting, 10X Genomics, sequencing and analysis of the clones. Z.S. performed bioinformatic analyses on 10X next-generation sequencing data and antibody repertoire. J.Y. cloned, expressed, and purified the mAbs. L.L. and Q.W. performed the epitope mapping and binding experiments. P.W. conducted the pseudovirus neutralization assays and M.S.N and Y.H.performed infectious SARS-CoV-2 neutralization assays. M.R., G.C., J.B., J.G., and L.S. carried out the cryo-EM studies. J.F.-W.C., Z.C., and K.-Y.Y. were responsible for the hamster experiment. Y.L. helped with project management. T.Z. and P.D.K. provided key reagents for the study, and P.D.K. contributed to the analysis and discussion of the data. L.L., P.W., M.S.N., J.Y., Y.H., Z.S., M.R., Q.W., L.S., and D.D.H. analysed the results, and D.D.H. wrote the manuscript, with contributions from each author. J.G.S. provided valuable suggestions.

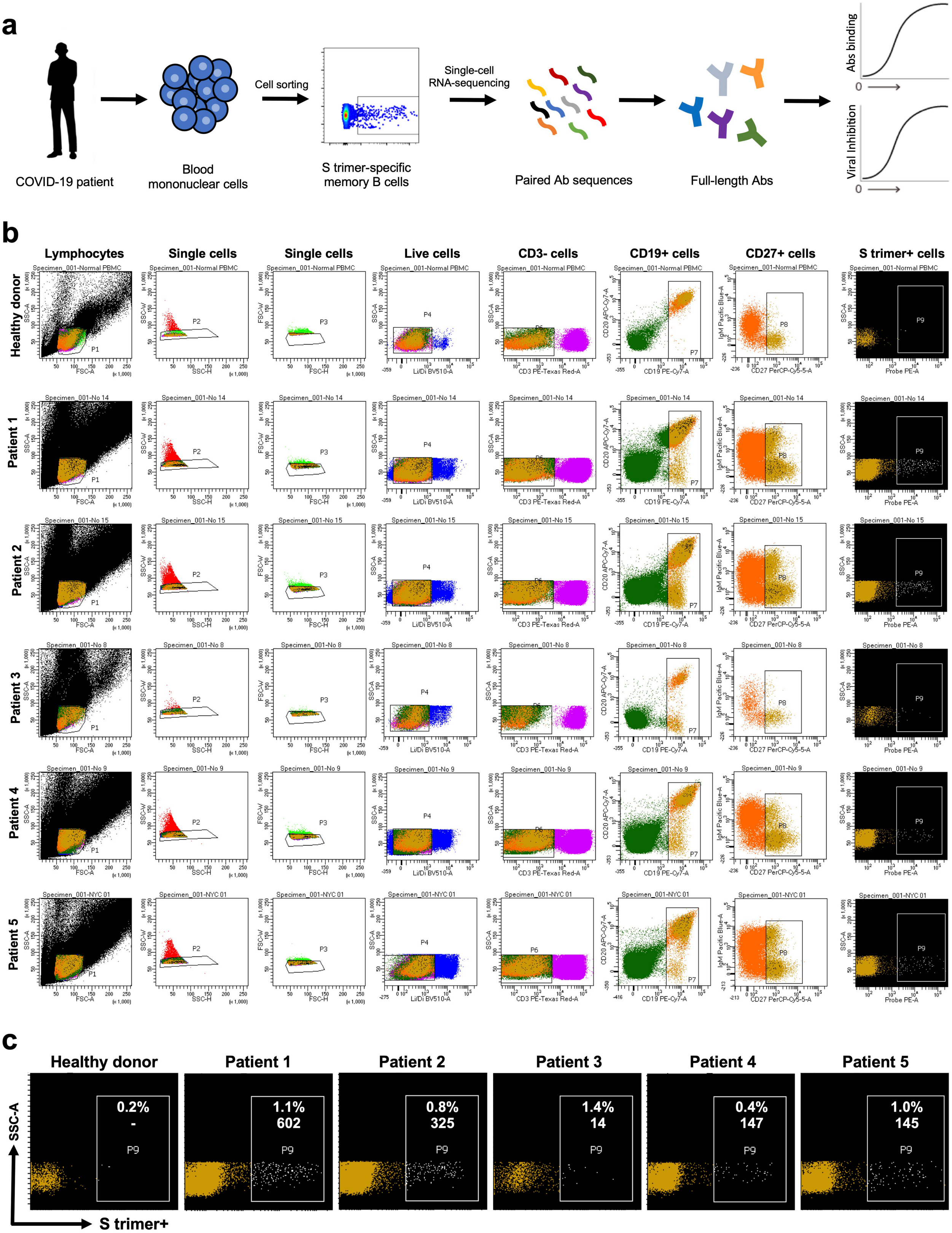
Competing interests A provisional patent application has been filed for the monoclonal antibodies described in the manuscript. L.L. and D.D.H. are inventors.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2571-7.

Correspondence and requests for materials should be addressed to Y.H., L.S. or D.D.H.

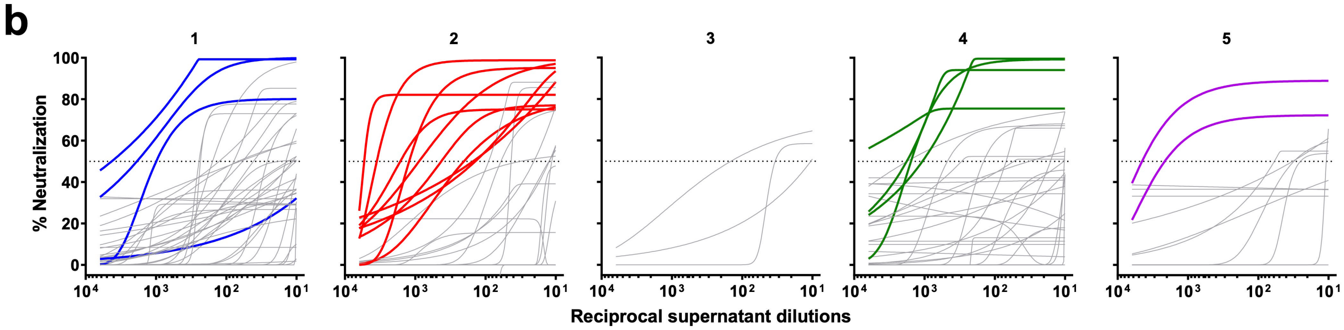
Peer review information *Nature* thanks Antonio Lanzavecchia and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



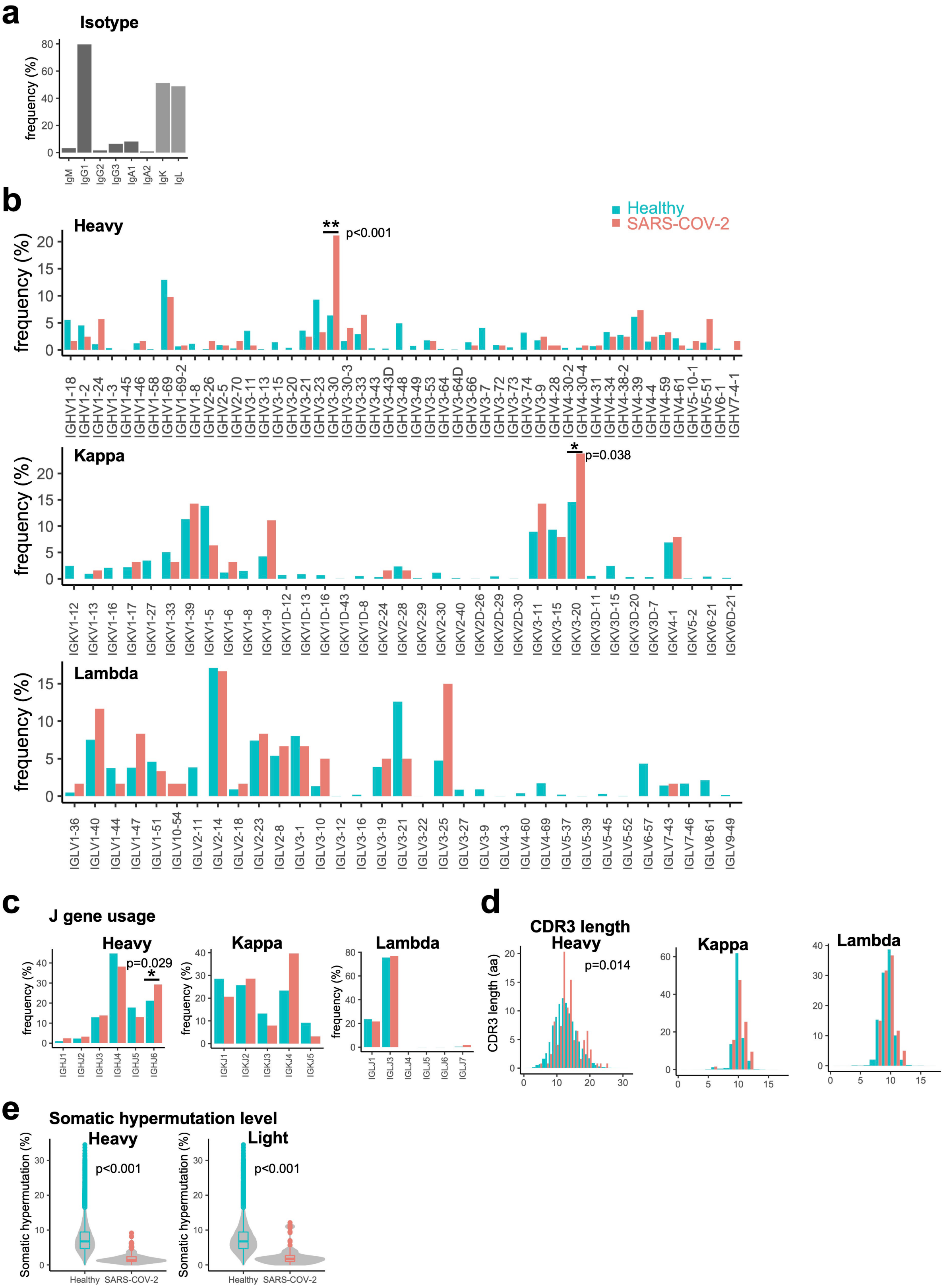
Extended Data Fig. 1 | SARS-CoV-2 S trimer-specific antibody isolation strategy. **a**, Schema for isolating of S trimer-specific mAbs from memory B cells in the blood of infected patients. **b**, Sorting results on the isolation of S trimer-specific memory B cells using flow cytometry. **c**, Magnified representation of the panel of S trimer-positive memory B cells for each patient. Inset numbers indicate the absolute number and the percentage of S trimer-specific memory B cells isolated from each case.

a

	Abs tested	Binding Abs			Neutralizing Abs	Neutralizing Abs
		S trimer	RBD	Non-RBD	Pseudovirus	Live virus
Total	252	121	38	83	61	41
Patient 1	100	45	19	26	19	11
Patient 2	54	29	12	17	18	18
Patient 3	6	2	0	2	3	0
Patient 4	44	32	7	25	14	6
Patient 5	48	13	0	13	7	6

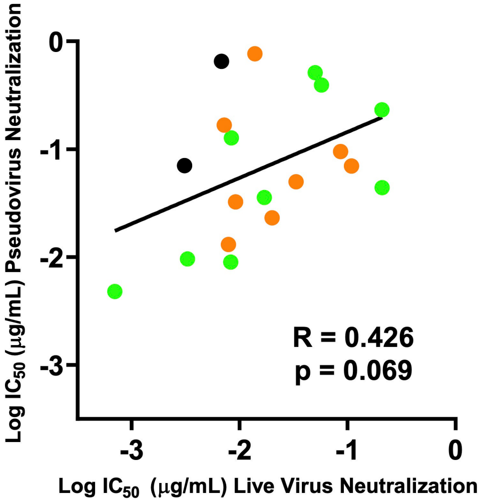


Extended Data Fig. 2 | Summary of mAb screening of transfection supernatants. a, Numbers of binding and neutralizing antibodies from patients 1 to 5. **b**, The best-fit pseudovirus neutralization curves for 130 samples that were positive in at least one of the screens shown in Fig. 1b. The 18 transfection supernatants that showed evidently better potency are highlighted in colours, while others with non-neutralizing or weakly neutralizing activities are shown in grey. One additional supernatant (Patient 1) that was initially missed in the pseudovirus screen but later found to be a potent neutralizing mAb (1-87) is also highlighted.

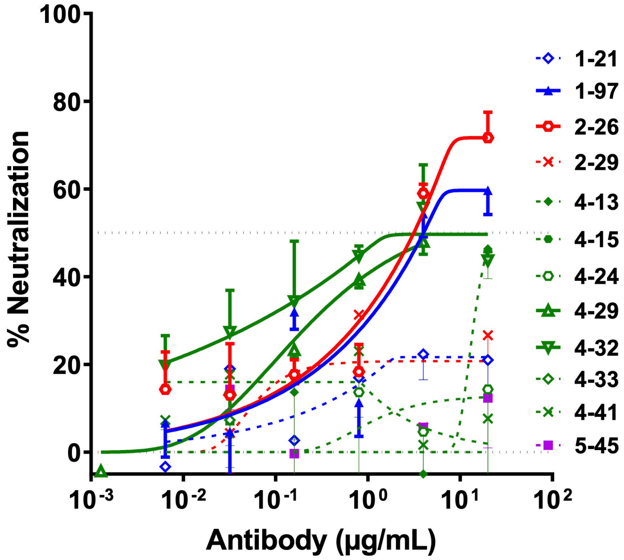


Extended Data Fig. 3 | See next page for caption.

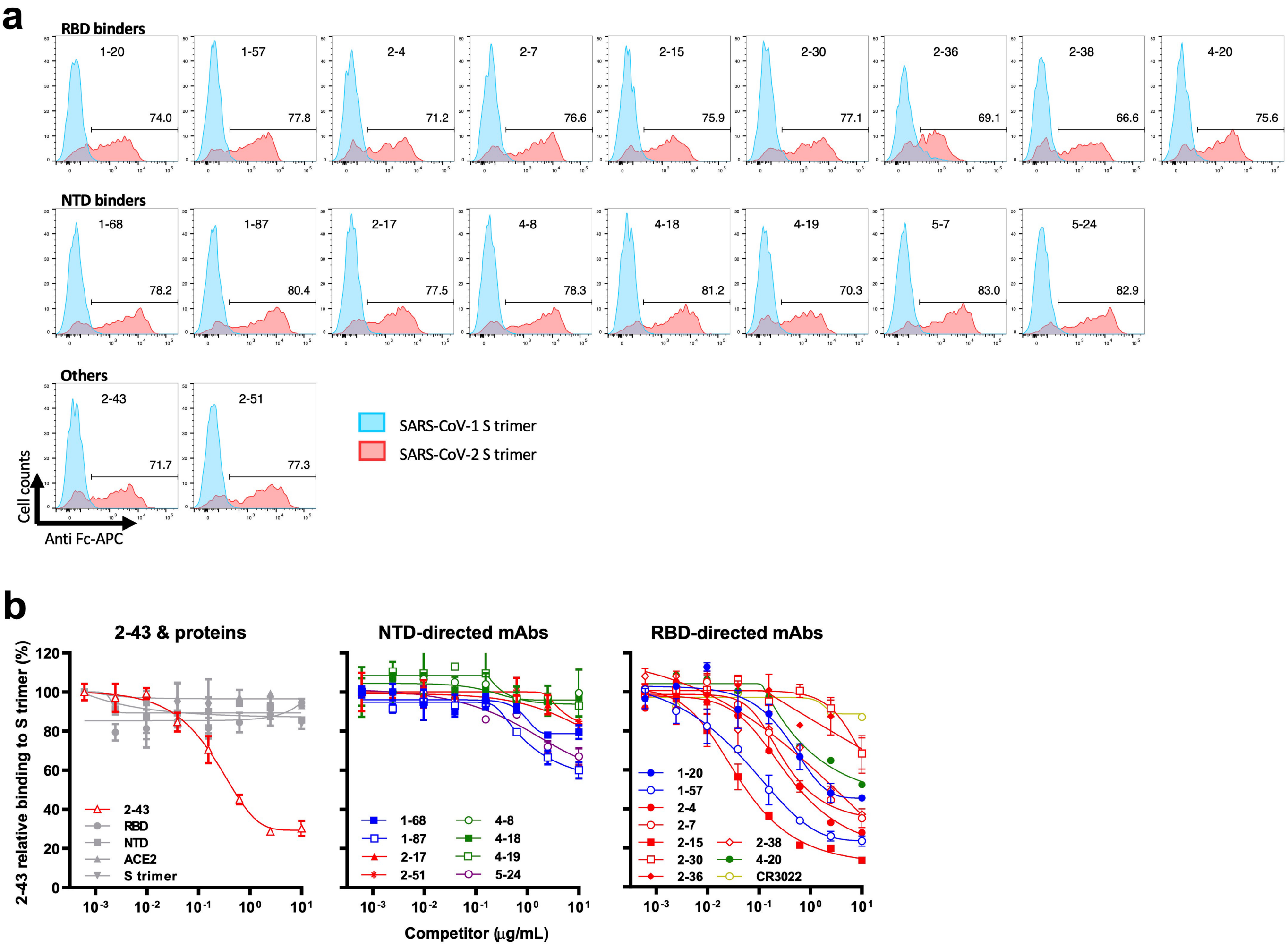
Extended Data Fig. 3 | Genetic features of SARS-CoV-2-specific antibody repertoire. a, 108 of the 123 antigen-specific antibodies are from IgG isotype. The kappa and lambda light chains are comparably used. **b**, Compared to IgG repertoires of healthy human donors (17,243, 27,575, and 20,889 transcripts for heavy, kappa, and lambda chains respectively), IGHV3-30 (antigen-specific $n = 26$ and healthy donor $n = 1117$) and IGKV3-20 genes (antigen-specific $n = 15$ and healthy donor $n = 4{,}071$) are over-represented in heavy and light chain repertoires respectively (P values are 6.415×10^{-11} and 0.04332 respectively, χ^2-test with 1 degree of freedom). We did not test the enrichment of other genes because the numbers of antigen-specific antibodies are less than 15. **c**, The usage of IGHJ6 gene (antigen-specific $n = 36$ and healthy donor $n = 3646$) was significantly higher in antigen-specific antibodies (χ^2-test with 1 degree of freedom, $P = 0.02807$). **d**, The CDRH3 length of antigen-specific antibodies is significantly longer than in healthy donors (two-sided Kolmogorov–Smirnov test, $P = 0.014$). **e**, For both heavy and light chains, the V region nucleotide somatic hypermutation levels are significantly lower than in antibodies of healthy donors (two-sided Kolmogorov–Smirnov test, $P < 2.2 \times 10^{-16}$ for both heavy and light chains). For the boxplots, the middle lines are medians. The lower and upper hinges correspond to the first and third quartiles respectively. The upper whisker extends to values no larger than 1.5× IQR (the interquartile range or distance between the first and third quartiles) from the hinge. The lower whisker extends to values no smaller than 1.5× IQR from the hinge. Data points beyond the whiskers were plotted as outliers using dots.



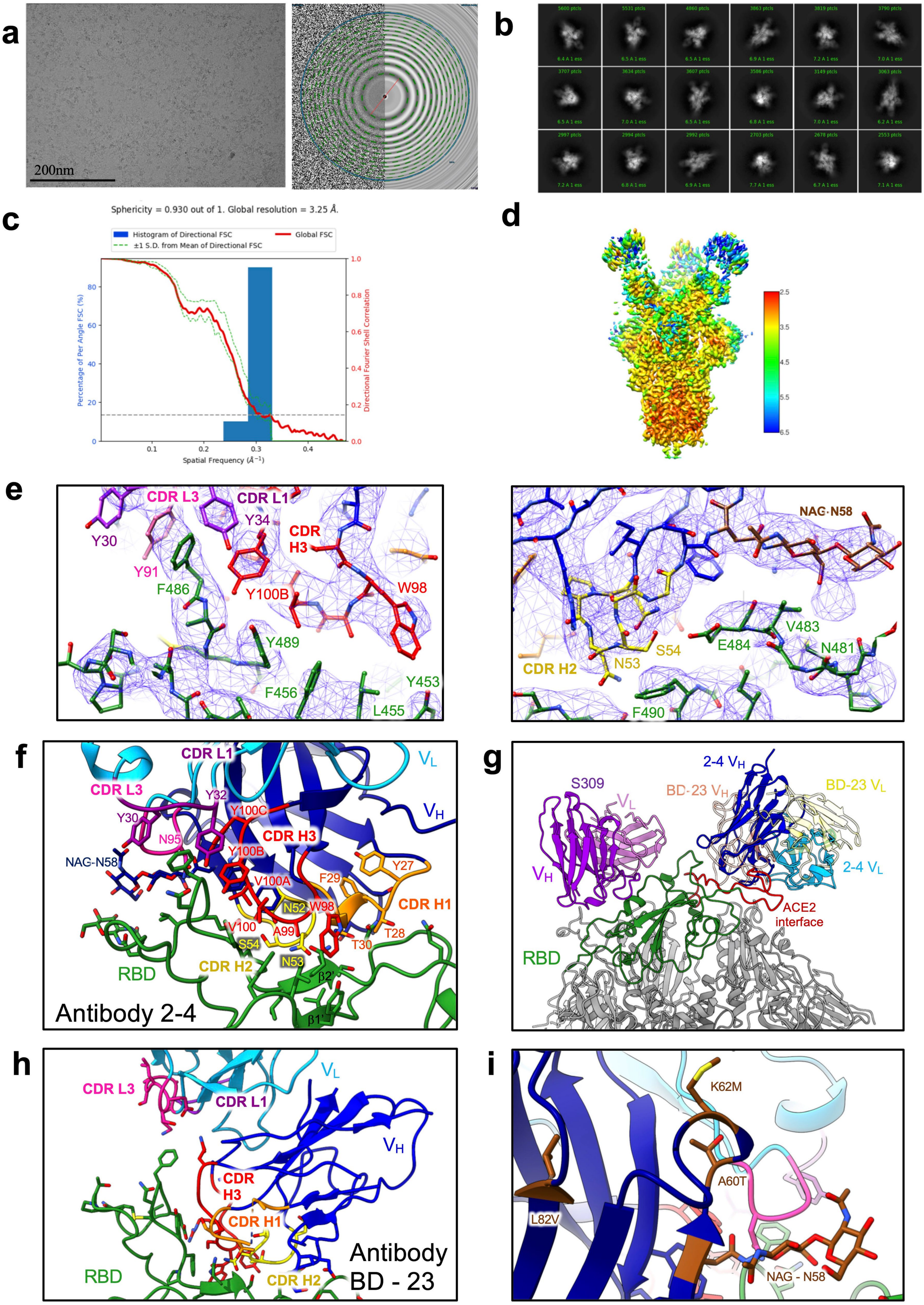
Extended Data Fig. 4 | Correlation of neutralizing antibody titres of the top 19 mAbs in the live SARS-CoV-2 assay versus the pseudovirus assay. Green circles represent RBD-directed antibodies; orange circles represent NTD-directed antibodies; and black circles represent antibodies in the 'Others' category. The Pearson correlation coefficient (*R*) and the *p* value were calculated using GraphPad Prism. Experiments were performed in triplicates for all mAbs tested.



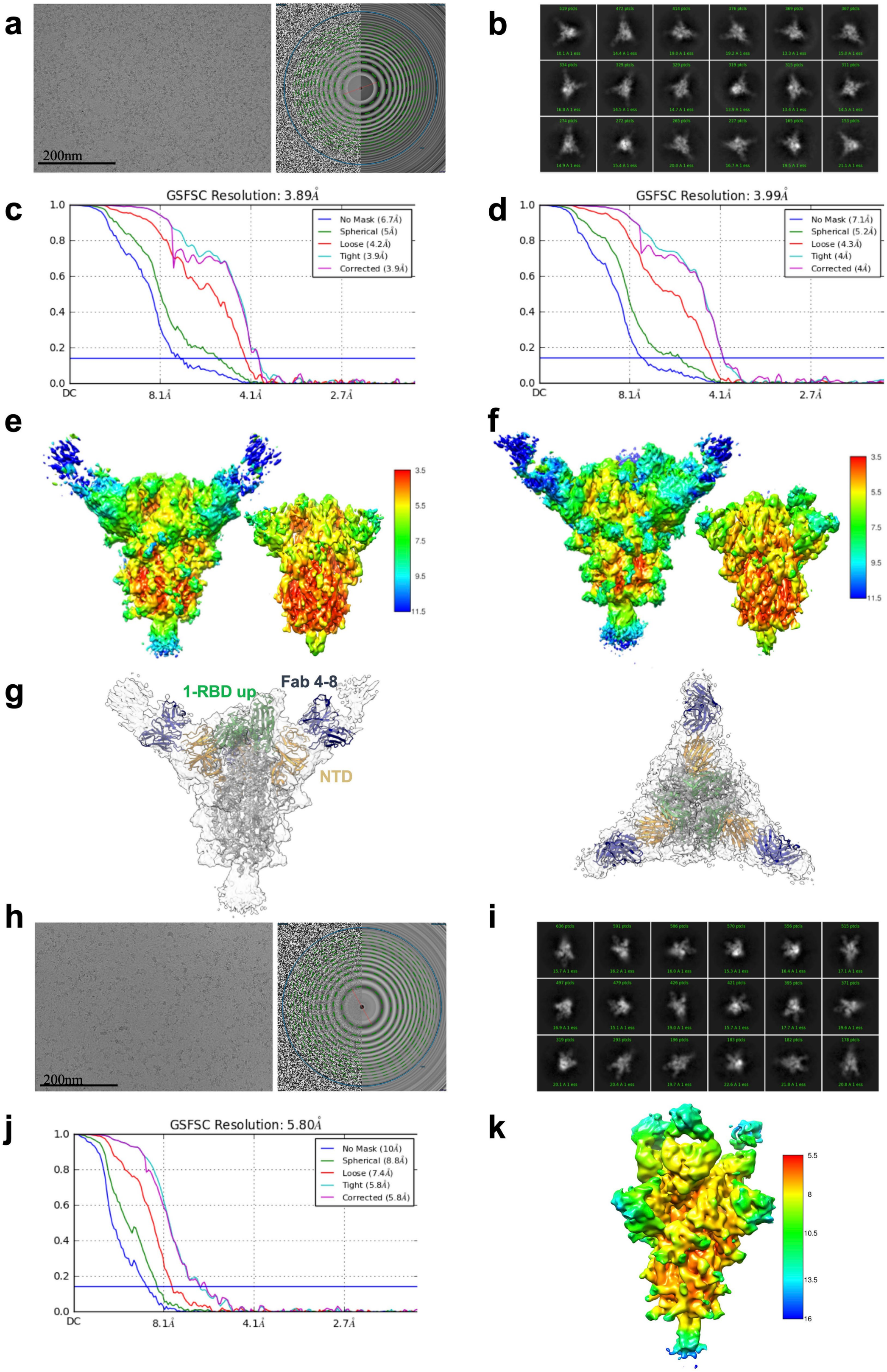
Extended Data Fig. 5 | The pseudovirus neutralization profiles for 12 purified mAbs that strongly bound the S trimer but with weak or no virus-neutralizing activities. The four mAbs with weak neutralizing activities against SARS-CoV-2 pseudovirus are shown in sold lines, and the remaining 8 non-neutralizing mAbs are shown in dashed lines.



Extended Data Fig. 6 | Cell-surface staining with antibodies. a, Antibody binding to the SARS-CoV-1 (blue) and SARS-CoV-2 (red) spike proteins expressed on the cell surface. Expi293 cells were co-transfected with GFP and full-length SARS-CoV-1 or SARS-CoV-2 spike genes. After 48 h, antibody binding to spike protein in the GFP-positive cells was detected by flow cytometer. The data show all antibodies tested were able to recognize the wildtype SARS-CoV-2 spike protein but not SARS-CoV-1 spike protein. **b**, Monoclonal Ab 2-43 bound to S trimer expressed on Expi293 cell surface can be competed out by mAbs directed against RBD but only minimally by mAbs to the NTD region. Shown are representative data from three independent experiments.



Extended Data Fig. 7 | Cryo-EM analysis of antibody 2-4 in complex with the S trimer. **a**, Representative micrograph and CTF of the micrograph. 8,324 micrographs were collected in total. **b**, Representative 2D class averages. **c**, Resolution of the consensus map with C3 symmetry as calculated by 3DFSC. **d**, The local resolution of the full map as calculated by cryoSPARC at an FSC cutoff of 0.5. **e**, Representative density of the Fab 2-4 (blue) and RBD (green) interface, showing interactions of CDR H3 in red, L1 in magenta, and L3 in light magenta (left), along with CDR H2 and the N-linked glycosylation added by SHM at ASN58 (right). **f**, Fab 2-4 binding interface with RBD. V_H is shown in blue, V_L in light blue, with CDRs H1 in orange, H2 in yellow, H3 in red, L1 in magenta, and L3 in light magenta. **g**, Positions of antibodies 2-4, S309[8], and BD-23[9] on the trimeric CoV-2 spike. **h**, Antibody BD-23[9] in complex with S trimer. **i**, Somatic hypermutations found only in the antibody 2-4 heavy chain, shown in brown. The mutation A60T creates an NxT sequence leading to N58 glycosylation.



Extended Data Fig. 8 | Cryo-EM data processing for antibodies 4-8 and 2-43 in complex with S trimer. **a**, Representative 4-8 micrograph and CTF of the micrograph. 3,153 micrographs were collected in total. **b**, Representative 2D class averages. **c**, Resolution of the spike in the RBD down conformation in complex with Fab 4-8. **d**, Resolution of the spike in the RBD up conformation in complex with Fab 4-8. **e**, Local resolution of the spike in the RBD down conformation in complex with Fab 4-8 at an FSC cutoff of 0.5, with two thresholds shown. **f**, Local resolution of the spike in the RBD up conformation in complex with Fab 4-8 at an FSC cutoff of 0.5, with two thresholds shown. **g**, Although the map was reconstructed at 4.0Å resolution, density for 4-8 Fab is poor due to molecular motion. A rigid body fit with SARS-CoV-2 spike and an antibody variable domain model is shown. **h**–**k**, Cryo-EM data processing for antibody 2-43 in complex with the S trimer. **h**, Representative 2-43 micrograph and CTF of the micrograph. **i**, Representative 2D class averages. **j**, Resolution of Fab 2-43 in complex with S trimer. **k**, The local resolution of the full map as calculated by cryoSPARC at an FSC cutoff of 0.5.

Extended Data Table 1 | Patient information

Patient	Age	Sex & Race	Days from symptom onset to:	Biomarker	Complications	Outcome
1	57	Female, Hispanic	Admission: 7 MV: 12 Ab isolation: 18	hsCRP = 208 mg/L ESR = 58 mm/hr IL-6 = 23 pg/mL Ferritin = 766 ng/mL D-dimer = 3.4 μg/mL FEU	ARDS	Discharged on day 30
2	71	Female, Hispanic	Admission: 20 MV: 20 Ab isolation: 29	hsCRP = 33 mg/L ESR > 130 mm/hr IL-6 = 13 pg/mL Ferritin = 425 ng/mL D-dimer = 5.7 μg/mL FEU	ARDS Ventilator associated pneumonia	Discharged on day 45
3	61	Male, White	Admission: 10 MV: 10 Ab isolation: 21	hsCRP = 51 mg/L ESR = 57 mm/hr IL-6 > 315 pg/mL Ferritin = 3,238 ng/mL D-dimer = 7.4 μg/mL FEU	ARDS Acute kidney injury (hemodialysis) Sepsis	Death on day 28
4	51	Male, Black	Admission: 7 MV: 10 Ab isolation: 25	hsCRP = 88 mg/L ESR = 110 mm/hr IL-6 = 77 pg/mL Ferritin = 510 ng/mL D-dimer = 13.4 μg/mL FEU	ARDS Acute kidney injury (no hemodialysis) Ventilator associated pneumonia	Discharged on day 51
5	50	Male, White	Admission: 5 MV: 7 Ab isolation: 32	hsCRP = 2 mg/L ESR = 63 mm/hr	ARDS Neuropathy	Discharged on day 27

ARDS, acute respiratory distress syndrome; MV, mechanical ventilation; hsCRP, high sensitivity C-reactive protein, ULN >10 mg/l; ESR, erythrocyte sedimentation rate, ULN = 20 mm/h; Interleukin 6, ULN = 5 pg/ml; Ferritin, ULN = 150 ng/ml; D-dimer quantitative ULN = 0.8 ug/ml FEU.

Extended Data Table 2 | Cryo-EM data collection, refinement, and validation statistics

	SARS-CoV-2 spike with Fab 2-4 (EMDB-22156) (PDB 6XEY)	SARS-CoV-2 spike RBD up with Fab 4-8 (EMDB-22158)	SARS-CoV-2 spike RBD down with Fab 4-8 (EMDB-22159)	SARS-CoV-2 spike with Fab 2-43 (EMDB-22275)
Data collection and processing				
Magnification	81,000	81,000	81,000	81,000
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	51.30	51.30	51.30	51.69
Defocus range (μm)	-0.4 to -3.5	-0.4 to -3.5	-0.4 to -3.5	-0.4 to -3.5
Pixel size (Å)	1.058	1.058	1.058	1.058
Symmetry imposed	C3	C1	C3	C1
Initial particle images (no.)	556,983	256,848	256,848	55,161
Final particle images (no.)	83,927	105,278	47,555	10,068
Map resolution (Å)	3.25	4.0	3.9	5.8
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	406.3-3.25	406.3-4.0	406.3-3.9	406.3-5.8
Refinement				
Initial model used (PDB code)	6VSB			
Model resolution (Å)	3.7			
FSC threshold	0.5			
Model resolution range (Å)	406.3-3.25			
Map sharpening *B* factor (Å^2)	-97.5			
Model composition				
Non-hydrogen atoms	28,482			
Protein residues	3788			
Ligands	63			
B factors (Å^2)				
Protein	54.35			
Ligand	73.91			
R.m.s. deviations				
Bond lengths (Å)	0.005			
Bond angles (°)	0.810			
Validation				
MolProbity score	1.51			
Clashscore	3.59			
Poor rotamers (%)	0.22			
Ramachandran plot				
Favored (%)	94.75			
Allowed (%)	5.17			
Disallowed (%)	0.08			

nature research

Corresponding author(s): David D. Ho

Last updated by author(s): Jul 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cryo-EM data was collected using Leginon 3.4.beta. Sequencing of memory B cell clones done using Illumina NextSeq 500. Cell sorting was performed on FACSDiva version 8.0.1.
Data analysis	Cryo-EM data was processed using cryoSPARC v2.14.2, MotionCor2, Topaz v0.2.4, 3DFSC v3.0, UCSF Chimera v1.13.1, ChimeraX v0.93, ISOLDE v1.0b5, Phenix v1.18, and COOT v0.8.9.2. Next-generation sequencing data of antibody repertoires were processed using Cell ranger v3.1.0, SONAR V1, BLAST v2.2.25, CLUSTALO1.2.3, and USEARCH v9.2.64. FlowJo 10.4 was used for analyzing FACS data. For 10X Genomics; cellranger 3.1.0 for BCL to FASTQ conversion, and gene counting was used. GraphPad Prism 8 was used for plotting data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We confirm we have deposited the sequencing dataset into Genbank and they will become available publicly within two business day. Once the accession numbers are assigned, we will add the data availability statement in the manuscript and here. Healthy donor antibody repertoires were from previous study with SRA ID PRJNA336331. The following data availability statement will be included in the final version of the manuscript: "The 19 neutralizing antibodies were deposited to Genbank with accession numbers: ACXXXXXXXX. Coordinates for the antibody 2-4 complex are deposited in the Protein Data Bank as PDB 6XEY. Cryo-EM maps and data are deposited in EMDB with deposition codes EMDB-22156 for antibody 2-4, EMDB-22158 and EMDB-22159 for antibody 4-8, and EMDB-22275 for antibody 2-43. These data are used in Fig. 4 and Extended Data Figs. 7, 8, 9, 10, and 11."

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	40 patients detected positive for SARS-CoV-2 using diagnostic RT-PCR tests were used for screening of neutralization abilities of their plasma samples. Based on the neutralization profile of the plasma, patients with most potent plasma were downselected for sorting of the memory B-cells and antibody isolation and cloning. The sample size is appropriate within technical capability to downselect multiple patients with potent neutralizing plasma.
Data exclusions	None
Replication	All experiments were performed and verified in multiple replicates as indicated in their methods/figure legends of the manuscript.
Randomization	All samples were selected for their ability to produce neutralization antibodies and all PBMCs were randomly processed from the 5 patients with potent neutralization of the plasma using baits specific for their ability to measure neutralization (SARS-CoV-2 S trimer). The screens for the binding and neutralization assays were also performed without any bias for selection and efficacy determined solely by the potency of the individual clones/antibodies.
Blinding	Blinded scoring of the neutralization of SARS-CoV-2 virus associated cytopathic effects were performed and average of the scores was converted to percentage of the neutralization. The results were plotted as mean +/- SEM. All other experiments in the study were predesigned with the hypothesis and strategies were laid out so as to use instruments that were calibrated to report the data. This feature led to the non-relevance of blinding for any of those experiments. Experiments were validated using technical and/or biological replicates in all cases.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	For S trimer-specific B cells sorting and single-cell BCR sequencing, anti-human CD3 PE-CF594 (BD Biosciences, Cat.562406, Clone SP34-2, Lot.9325656, 1:20 dilution), anti-human CD19 397 PE-Cy7 (Biolegend, Cat.302216, Clone HIB19, Lot.B276834, 1:20 dilution), anti-human CD20 APC-Cy7 (Biolegend, Cat.302314, Clone 2H7, Lot.B288789, 1:20 dilution), anti-human IgM V450 (BD Biosciences, Cat.561286, Clone G20-127, Lot.9003910, 1:20 dilution), anti-human CD27 PerCP-Cy5.5 (BD Biosciences, Cat.560612, Clone M-T271, Lot.9283016, 1:20 dilution), anti-His PE (Biolegend, Cat.362603, Clone J095G46, Lot.B269138, 1:20 dilution), Human Hashtag 3 (Biolegend, Cat.394665, Clone LNH-94,Lot.B282244, 1:20 dilution). For epitope mapping by ELISA, anti-human IgG (Jackson ImmunoResearch, Cat. 109-035-003, Polyclonal, Lot.146269, 1: 10,000 dilution), Streptavidin-APC (Biolegend, Cat.405243, Lot.B266052, 1: 2,000 dilution), Avidin-HRP (Invitrogen, Cat.18-4100-51, Lot.2197902, 1: 500 dilutions), anti-Strep-HRP (Strep-TagII -HRP, EMD Millipore, Cat.71591, Lot.3393843, 1: 2,000 dilution).
Validation	All validations are available from the commercial website under the validation sheet link for the catalogued item. 1. Anti-human CD3 PE-CF594 (BD Biosciences, Cat # 562406), https://www.bdbiosciences.com/eu/reagents/research/antibodies-buffers/immunology-reagents/anti-non-human-primate-antibodies/cell-surface-antigens/pe-cf594-mouse-anti-human-cd3-sp34-2/

p/562406
2. Anti-human CD19 397 PE-Cy7 (Biolegend, Cat# 302216), https://www.biolegend.com/en-us/products/pe-cyanine7-anti-human-cd19-antibody-1911
3. Anti-human CD20 APC-Cy7 (Biolegend, Cat# 302314), https://www.biolegend.com/en-us/products/apc-cyanine7-anti-human-cd20-antibody-1901
4. Anti-human IgM V450 (BD Biosciences, Cat # 561286), https://www.bdbiosciences.com/eu/applications/research/b-cell-research/immunoglobulins/human/v450-mouse-anti-human-igm-g20-127/p/561286
5. Anti-human CD27 PerCP-Cy5.5 (BD Biosciences, Cat# 560612), https://www.bdbiosciences.com/eu/applications/research/b-cell-research/surface-markers/human/percp-cy55-mouse-anti-human-cd27-m-t271/p/560612
6. Human Hashtag 3 (Biolegend, Cat # 394665), https://www.biolegend.com/en-us/products/totalseq-c0253-anti-human-hashtag-3-antibody-17164
7. Anti-His PE (Biolegend, Cat# 362603), https://www.biolegend.com/en-us/products/pe-anti-his-tag-antibody-9861
8. Anti-human IgG (Jackson ImmunoResearch, Cat# 109-035-003), https://www.jacksonimmuno.com/catalog/products/109-035-003
9. Streptavidin-APC (Biolegend, Cat# 405243), https://www.biolegend.com/en-us/products/apc-streptavidin-high-concentration-10081
10. Avidin-HRP (Invitrogen, Cat# 18-4100-51), https://www.thermofisher.com/order/catalog/product/18-4100-51#/18-4100-5
11. Anti-Strep-HRP (Strep-TagII –HRP, EMD Millipore, Cat# 71591), https://www.emdmillipore.com/US/en/product/StrepTag-II-Antibody-HRP-Conjugate,EMD_BIO-71591

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Vero-E6 (ATCC), Expi293 (Thermofisher), 293T (ATCC)
Authentication	Obtained from authenticated vendors. Cells were recovered as healthy logarithmically growing cells within 4 to 7 days after thawing. Viability was measured and found to be >90%.
Mycoplasma contamination	Mycoplasma is negative (Detected mycoplasma contamination using Mycoplasma PCR ELISA ,Sigma,catalog number is 11663925910)
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in the study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Eligibility criteria include: (1) greater than age 18 (inclusive) (2) confirmed COVID-19 infection by a FDA- approved molecular based assay (including those under emergency use authorization) of respiratory or blood specimens; (3) If symptomatic with COVID-19, must have evidence of improvement of symptoms and a duration of at least 4 weeks from the onset of symptoms to day of enrollment; (4) If asymptomatic, must have a duration of at least 4 weeks from first positive molecular based COVID-19 assay to day of enrollment. Among the 40 participants enrolled in this study, the mean age was 50 (20-84) and 53% were male. Among those with race/ethnicity information, 21% were Black/African American, 38% Latinx, 3% Asian, and 38% non-Hispanic white.
Recruitment	This is a prospective study to enroll participants who have recovered from coronavirus disease (COVID-19) for the purpose of obtaining blood specimens to isolate monoclonal antibodies against SARS-CoV2 that can be developed into preventive or therapeutic agents. Potential participants were referred by health care providers from within the Columbia University Irving Medical Center/New York Presbyterian Hospital system and from outside institutions. Potential participants were contacted by study staff and informed consent signed prior to performance of study procedures. All participants with severe COVID-19 were recruited during or after prolonged hospitalization at a single medical center in New York City, while participants with mild COVID-19 were self-referred through online recruitment. All participants were recruited in March and April, 2020 during the early stages of the epidemic in New York. These factors may impact the generalizability of our findings.
Ethics oversight	This protocol, “Acquiring convalescent specimens to isolate and identify potent monoclonal antibodies against COVID-19” (AAAS9517) was approved by the Columbia University Institutional Review Board. Informed consent was obtained from all participants or surrogates. This statement is added to the manuscript.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	NCT04342195
Study protocol	The protocol "Acquiring convalescent specimens to isolate and identify potent monoclonal antibodies against COVID-19" is accessible by sending request to Dr. Michael Yin <mty4@cumc.columbia.edu>.

Data collection	The study protocol was approved on 3/13/2020 and the last participant enrolled for this analysis was on 4/7/2020. All data were collected at Columbia University Irving Medical Center, New York NY. Recruitment and data collection occurred between 3/25/2020 and 4/7/2020.
Outcomes	The primary outcome for the clinical study was the SARS-CoV-2 antibody response as measured by the S-trimer and nucleocapsid ELISA and pseudovirus assays.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Peripheral blood mononuclear cells from five patients and one healthy donor were stained with LIVE/DEAD™ Fixable Yellow Dead Cell Stain Kit (Invitrogen) at ambient temperature for 20 mins, followed by washing with RPMI-1640 complete medium and incubation with 10 µg/mL of S trimer at 4°C for 45 mins. Afterwards, the cells were washed again and incubated with a cocktail of flow cytometry and hashtag antibodies, containing CD3 PE-CF594 (BD Biosciences), CD19 PE-Cy7 (Biolegend), CD20 APC-Cy7 (Biolegend), IgM V450 (BD Biosciences), CD27 PerCP Cy5.5 (BD Biosciences), anti-His PE (Biolegend), and human Hashtag 3 (Biolegend) at 4°C for 1hr. Stained cells were then washed, resuspended in RPMI-1640 complete medium and sorted for S trimer-specific memory B cells (CD3-CD19+CD27+S trimer+ live single lymphocytes).
Instrument	BD FACSAriaII (P69500149)
Software	FACSDiva version 8.0.1
Cell population abundance	S trimer bait positive cells were purified from the PBMCs of the 5 patients using the gating strategy used below. Purified trimer positive memory B cells were obtained from 5 patients and compared to healthy donor (negative control) as shown in extended data figure 1.
Gating strategy	As shown in Supplementary Figure 1b, sorting of the PBMC was performed in identical manner for all the samples including healthy donor. The summary of the gating is provided herewith: All PBMCs were initially gated using FSC-A and SSC-A gates for lymphocyte populations. The lymphocytes were gated using SSC-H and SSC-W initially followed by FSC-H and FSC-W to isolate the singlets in the population. The singlets was gated based on the fluorescence from the LIVE/DEAD™ Fixable Yellow Dead Cell Stain Kit for live cells. This step was followed by selecting for CD3- population by gating the SSC-A versus CD3-PE-CF594 stained population on the Texas Red channel. The negative population was gated for B-cells by first selecting for CD19 + cells followed by CD27+ cells on the respective fluorescent channels. The subsets of CD19+ cells were then selected for S-trimer bait positive by selecting for the cells bound to anti-Histag-PE on the trimer.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls

https://doi.org/10.1038/s41586-020-2550-z

Received: 20 May 2020

Accepted: 7 July 2020

Nina Le Bert[1,9], Anthony T. Tan[1,9], Kamini Kunasegaran[1], Christine Y. L. Tham[1], Morteza Hafezi[1], Adeline Chia[1], Melissa Hui Yen Chng[1], Meiyin Lin[1,2], Nicole Tan[1], Martin Linster[1], Wan Ni Chia[1], Mark I-Cheng Chen[3], Lin-Fa Wang[1], Eng Eong Ooi[1], Shirin Kalimuddin[4], Paul Anantharajah Tambyah[5,6], Jenny Guek-Hong Low[1,4], Yee-Joo Tan[2,7] & Antonio Bertoletti[1,8] ✉

Memory T cells induced by previous pathogens can shape susceptibility to, and the clinical severity of, subsequent infections[1]. Little is known about the presence in humans of pre-existing memory T cells that have the potential to recognize severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Here we studied T cell responses against the structural (nucleocapsid (N) protein) and non-structural (NSP7 and NSP13 of *ORF1*) regions of SARS-CoV-2 in individuals convalescing from coronavirus disease 2019 (COVID-19) ($n = 36$). In all of these individuals, we found CD4 and CD8 T cells that recognized multiple regions of the N protein. Next, we showed that patients ($n = 23$) who recovered from SARS (the disease associated with SARS-CoV infection) possess long-lasting memory T cells that are reactive to the N protein of SARS-CoV 17 years after the outbreak of SARS in 2003; these T cells displayed robust cross-reactivity to the N protein of SARS-CoV-2. We also detected SARS-CoV-2-specific T cells in individuals with no history of SARS, COVID-19 or contact with individuals who had SARS and/or COVID-19 ($n = 37$). SARS-CoV-2-specific T cells in uninfected donors exhibited a different pattern of immunodominance, and frequently targeted NSP7 and NSP13 as well as the N protein. Epitope characterization of NSP7-specific T cells showed the recognition of protein fragments that are conserved among animal betacoronaviruses but have low homology to 'common cold' human-associated coronaviruses. Thus, infection with betacoronaviruses induces multi-specific and long-lasting T cell immunity against the structural N protein. Understanding how pre-existing N- and ORF1-specific T cells that are present in the general population affect the susceptibility to and pathogenesis of SARS-CoV-2 infection is important for the management of the current COVID-19 pandemic.

SARS-CoV-2 is the cause of COVID-19[2]. This disease has been declared a pandemic by the World Health Organization (WHO), and is having severe effects on both individual lives and economies around the world. Infection with SARS-CoV-2 is characterized by a broad spectrum of clinical syndromes, which range from asymptomatic disease or mild influenza-like symptoms to severe pneumonia and acute respiratory distress syndrome[3].

It is common to observe the ability of a single virus to cause widely differing pathological manifestations in humans. This is often due to multiple contributing factors including the size of the viral inoculum, the genetic background of patients and the presence of concomitant pathological conditions. Moreover, an established adaptive immunity towards closely related viruses[4] or other microorganisms[5] can reduce susceptibility[6] or enhance disease severity[7].

SARS-CoV-2 belongs to the *Coronaviridae*, a family of large RNA viruses that infect many animal species. Six other coronaviruses are known to infect humans. Four of them are endemically transmitted[8] and cause the common cold (OC43, HKU1, 229E and NL63), while SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) have caused epidemics of severe pneumonia[9]. All of these coronaviruses trigger antibody and T cell responses in infected patients: however, antibody levels appear to wane faster than T cells. SARS-CoV-specific antibodies dropped below the limit of detection within 2 to 3 years[10], whereas SARS-CoV-specific memory T cells have been detected even 11 years after SARS[11]. As the sequences of selected structural and non-structural proteins are highly conserved among different coronaviruses (for example, NSP7 and NSP13 are 100% and 99% identical, respectively, between SARS-CoV-2, SARS-CoV and the bat-associated bat-SL-CoVZXC21[12]), we investigated whether cross-reactive SARS-CoV-2-specific T cells are present in individuals who resolved SARS-CoV, and compared the responses with those present in individuals who recovered from SARS-CoV-2 infection. We also

---

[1]Emerging Infectious Diseases Program, Duke-NUS Medical School, Singapore, Singapore. [2]Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore, Singapore. [3]National Centre of Infectious Diseases, Singapore, Singapore. [4]Department of Infectious Diseases, Singapore General Hospital, Singapore, Singapore. [5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [6]Division of Infectious Disease, University Medicine Cluster, National University Hospital, Singapore, Singapore. [7]Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [8]Singapore Immunology Network, A*STAR, Singapore, Singapore. [9]These authors contributed equally: Nina Le Bert, Anthony T. Tan. ✉e-mail: antonio@duke-nus.edu.sg

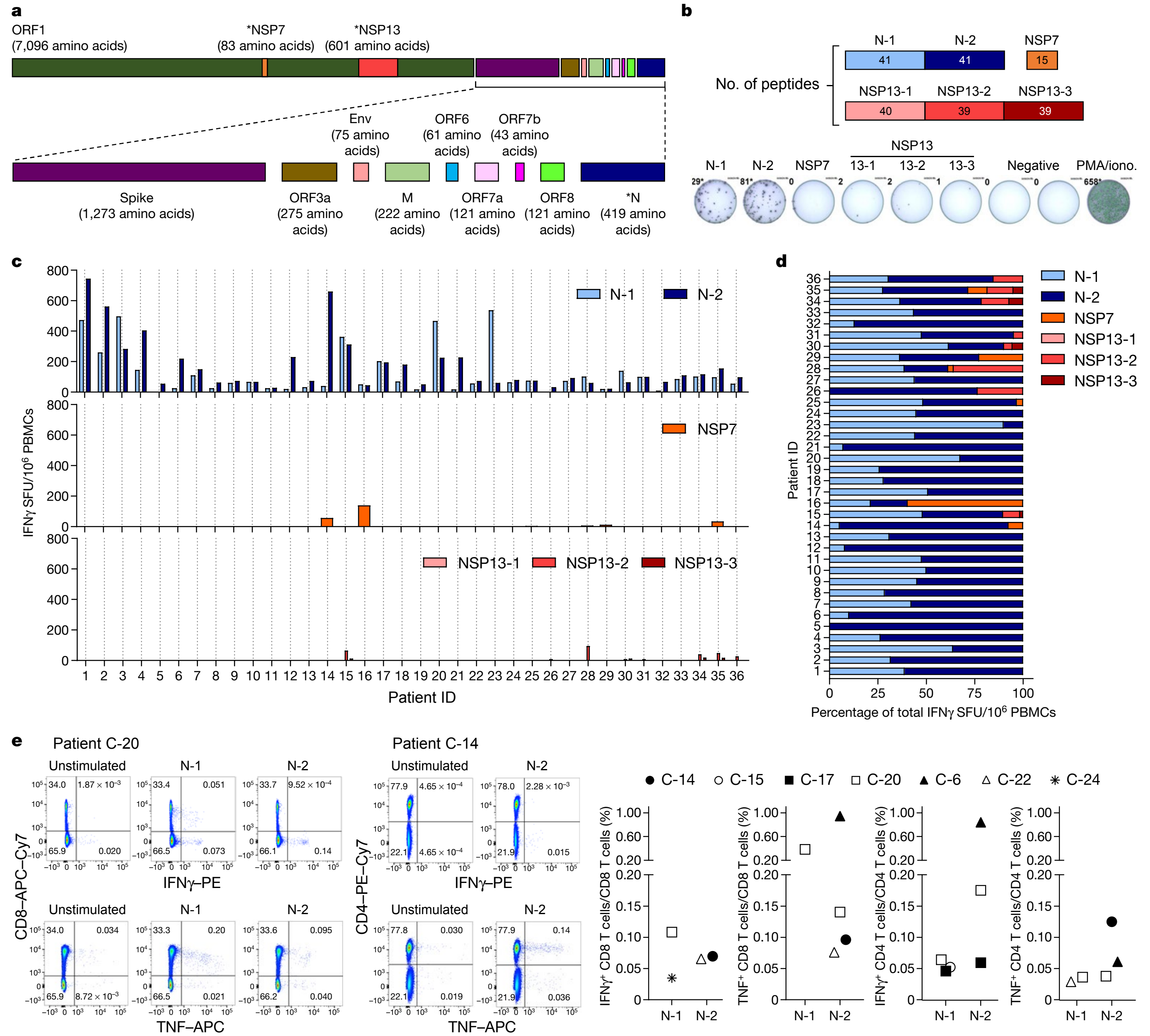


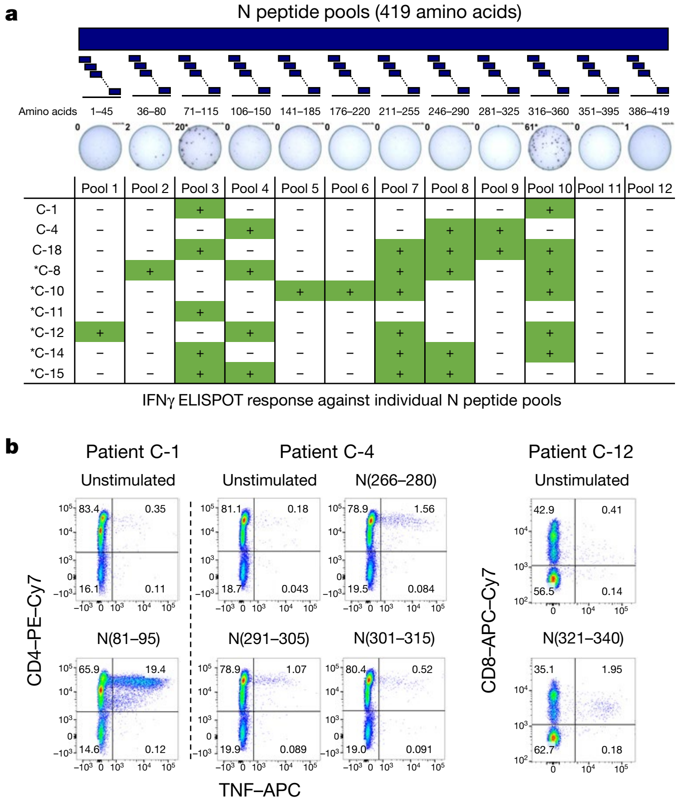
Fig. 1 | SARS-CoV-2-specific responses in patients recovered from COVID-19. a, SARS-CoV-2 proteome organization; analysed proteins are marked by an asterisk. **b**, The 15-mer peptides, which overlapped by 10 amino acids, comprising the N protein, NSP7 and NSP13 were split into 6 pools covering the N protein (N-1, N-2), NSP7 and NSP13 (NSP13-1, NSP13-2, NSP13-3). **c**, PBMCs of patients who recovered from COVID-19 (*n* = 36) were stimulated with the peptide pools or with phorbol 12-myristate 13-acetate (PMA) and ionomycin (iono) as a positive control. The frequency of spot-forming units (SFU) of IFNγ-secreting cells is shown. **d**, The composition of the SARS-CoV-2 response in each individual is shown as a percentage of the total detected response. N-1, light blue; N-2, dark blue; NSP7, orange; NSP13-1, light red; NSP13-2, red; NSP13-3, dark red. **e**, PBMCs were stimulated with the peptide pools covering the N protein (N-1, N-2) for 5 h and analysed by intracellular cytokine staining. Dot plots show examples of patients (2 out of 7) that had CD4 and/or CD8 T cells that produced IFNγ and/or TNF in response to stimulation with N-1 and/or N-2 peptides. The percentage of SARS-CoV-2 N-peptide-reactive CD4 and CD8 T cells in *n* = 7 individuals are shown (unstimulated controls were subtracted for each response).

studied these T cells in individuals with no history of SARS or COVID-19 or of contact with patients with SARS-CoV-2. Collectively these individuals are hereafter referred to as individuals who were not exposed to SARS-CoV and SARS-CoV-2 (unexposed donors).

SARS-CoV-2-specific T cells in patients with COVID-19

SARS-CoV-2-specific T cells have just started to be characterized for patients with COVID-19[13,14] and their potential protective role has been inferred from studies of patients who recovered from SARS[15] and MERS[16]. To study SARS-CoV-2-specific T cells associated with viral clearance, we collected peripheral blood from 36 individuals after recovery from mild to severe COVID-19 (demographic, clinical and virological information is included in Extended Data Table 1) and studied the T cell response against selected structural (N) and non-structural proteins (NSP7 and NSP13 of ORF1) of the large SARS-CoV-2 proteome (Fig. 1a). We selected the N protein as it is one of the more-abundant structural proteins produced[17] and has a high degree of homology between different betacoranaviruses[18] (Extended Data Fig. 1).

NSP7 and NSP13 were selected for their complete homology between SARS-CoV, SARS-CoV-2 and other animal coronaviruses that belong to the betacoranavirus genus[12] (Extended Data Fig. 2), and because



	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5	Pool 6	Pool 7	Pool 8	Pool 9	Pool 10	Pool 11	Pool 12
C-1	–	–	+	–	–	–	–	–	–	+	–	–
C-4	–	–	–	+	–	–	–	+	+	–	–	–
C-18	–	–	+	–	–	–	+	+	+	+	–	–
*C-8	–	+	–	+	–	–	+	+	–	+	–	–
*C-10	–	–	–	–	+	+	+	–	–	+	–	–
*C-11	–	–	+	–	–	–	–	–	–	–	–	–
*C-12	+	–	–	+	–	–	+	–	–	+	–	–
*C-14	–	–	+	–	–	–	+	+	–	+	–	–
*C-15	–	–	+	+	–	–	+	+	–	–	–	–

IFNγ ELISPOT response against individual N peptide pools

Fig. 2 | SARS-CoV-2-specific T cells in COVID-19 convalescent individuals target multiple regions of the N protein. a, PBMCs of 9 individuals who recovered from COVID-19 were stimulated with 12 different pools of 7–8 N peptides. The table shows IFNγ ELISpot responses against the individual N peptide pools. The asterisk denotes responses detected after in vitro expansion. **b**, After in vitro cell expansion, a peptide pool matrix strategy was used. T cells that reacted to distinct peptides were identified by IFNγ ELISpot and confirmed by ICS. Representative dot plots of 3 out of 7 patients are shown.

they are representative of the ORF1a/b polyprotein that encodes the replicase–transcriptase complex[19]. This polyprotein is the first to be translated after infection with coronavirus and is essential for the subsequent transcription of the genomic and sub-genomic RNA species that encode the structural proteins[19]. We synthesized 216 15-mer peptides that overlapped by 10 amino acids and that covered the whole length of NSP7 (83 amino acids), NSP13 (601 amino acids) and N (422 amino acids) and split these peptides into five pools of approximately 40 peptides each (N-1, N-2, NSP13-1, NSP13-2 and NSP13-3) and a single pool of 15 peptides that spanned NSP7 (Fig. 1b). This unbiased method with overlapping peptides was used instead of bioinformatics selection of peptides, as the performance of such algorithms is often sub-optimal in Asian populations[20].

Peripheral blood mononuclear cells (PBMCs) of 36 patients who recovered from COVID-19 were stimulated for 18 h with the different peptide pools and virus-specific responses were analysed by interferon-γ (IFNγ) ELISpot assay. In all individuals tested (36 out of 36), we detected IFNγ spots after stimulation with the pools of synthetic peptides that covered the N protein (Fig. 1c, d). In nearly all individuals, N-specific responses could be identified against multiple regions of the protein: 34 out of 36 individuals showed reactivity against the region that comprised amino acids 1–215 (N-1) and 36 out of 36 individuals showed reactivity against the region comprising amino acids 206–419 (N-2). By contrast, responses to NSP7 and NSP13 peptide pools were detected at very low levels in 12 out of 36 COVID-19-convalescent individuals tested.

Direct ex vivo intracellular cytokine staining (ICS) was performed to confirm and define the N-specific IFNγ ELISpot response. Owing to their relative low frequency, N-specific T cells were more difficult to visualize by ICS than by ELISpot; however, a clear population of CD4 and/or CD8 T cells that produced IFNγ and/or TNF was detectable in seven out of nine analysed individuals (Fig. 1e and Extended Data Figs. 3, 4). Moreover, despite the small sample size, we could compare the frequency of SARS-CoV-2-specific IFNγ spots with the presence of virus-neutralizing antibodies, the duration of infection and disease severity and found no correlations (Extended Data Fig. 5). To confirm and further delineate the multi-specificity of the N-specific responses detected ex vivo in patients who recovered from COVID-19, we mapped the precise regions of the N protein that is able to activate IFNγ responses in nine individuals. We organized the 82 overlapping peptides that covered the entire N protein into small peptide pools (of 7–8 peptides) that were used to stimulate PBMCs either directly ex vivo or after an in vitro expansion protocol that has previously been used for patients with hepatitis B virus[21] or SARS[22]. A schematic representation of the peptide pools is shown in Fig. 2a. We found that 8 out of 9 patients who recovered from COVID-19 had PBMCs that recognized multiple regions of the N protein of SARS-CoV-2 (Fig. 2a). Notably, we then defined single peptides that were able to activate T cells in seven patients. Using a peptide matrix strategy[22], we first deconvolved the individual peptides that were responsible for the detected response by IFNγ ELISpot. Subsequently, we confirmed the identity of the single peptides by testing—using ICS—the ability of the peptides to activate

Table 1 | SARS-CoV-2-specific T cell epitopes

Participants	T cell phenotype	Protein (amino acid residues)	SARS-CoV-2 amino acid sequence	SARS-CoV amino acid sequence
C-1	CD4	N (81–95)	DDQIGYYRRATRRIR	DDQIGYYRRATRRVR
	CD8	N (321–340)	G**MEVTPSGTWL**TYTGAIKLD	G**MEVTPSGTWL**TYHGAIKLD
C-4	CD4	N (266–280)	KAYNVTQAFGRRGPE	KQYNVTQAFGRRGPE
	CD4	N (291–305)	LIRQGTDYKHWPQIA	LIRQGTDYKHWPQIA
	CD4	N (301–315)	WPQIAQFAPSASAFF	WPQIAQFAPSASAFF
C-8	CD4	N (51–65)	SWFTALTQHGKEDLK	SWFTALTQHGKEELR
	CD4	N (101–120)	**MKDLSPRWYFYYLG**TGPEAG	**MKELSPRWYFYYLG**TGPEAS
C-10	CD4 and CD8	N (321–340)	G**MEVTPSGTWL**TYTGAIKLD	G**MEVTPSGTWL**TYHGAIKLD
C-12	CD8	N (321–340)	G**MEVTPSGTWL**TYTGAIKLD	G**MEVTPSGTWL**TYHGAIKLD
C-15	CD4	N (101–120)	**MKDLSPRWYFYYLG**TGPEAG	**MKELSPRWYFYYLG**TGPEAS
C-16	CD4	NSP7 (21–35)	RVESSSKLWAQCVQL	RVESSSKLWAQCVQL

T cells that react with distinct peptides were identified by IFNγ ELISpot and confirmed by ICS. Previously described T cell epitopes for SARS-CoV are highlighted in bold; non-conserved amino acid residues between SARS-CoV and SARS-CoV-2 are underlined.

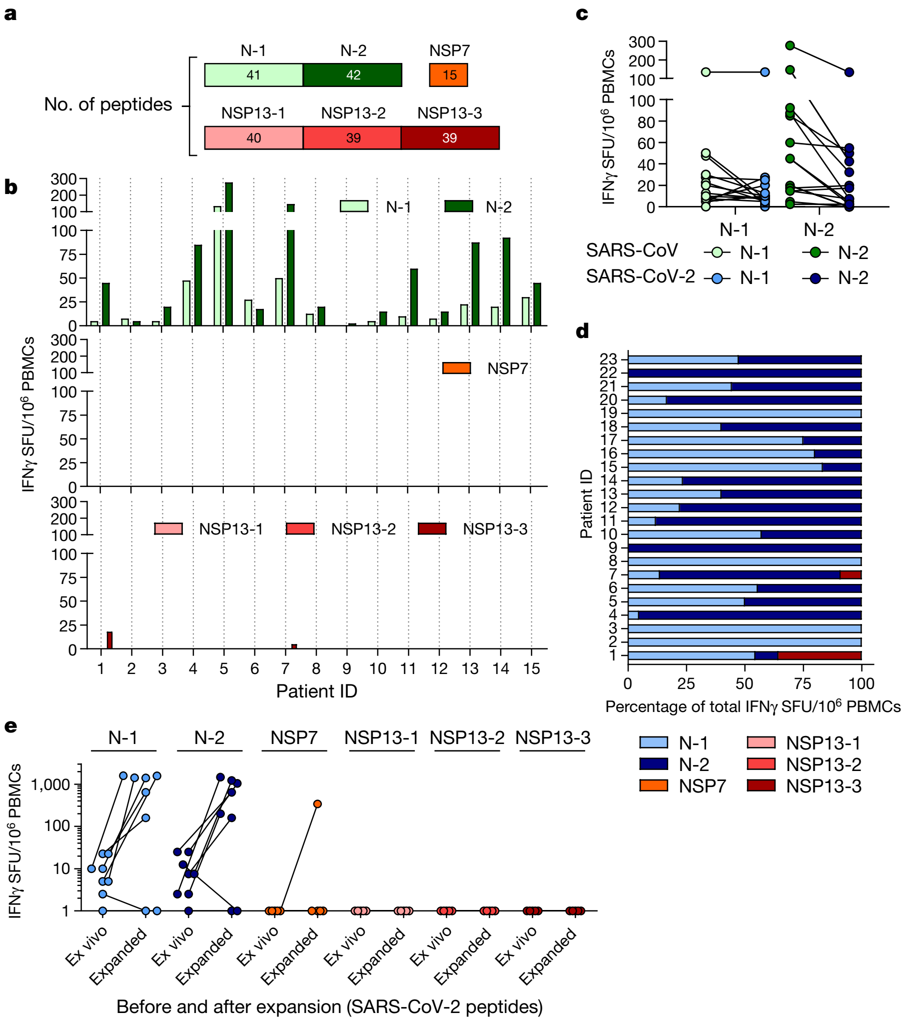


Fig. 3 | SARS-CoV-2 cross-reactive responses are present in patients who recovered from SARS. **a**, PBMCs isolated from 15 individuals who recovered from SARS 17 years ago were stimulated with SARS-CoV N, NSP7 and NSP13 peptide pools. **b**, Spot-forming units of IFNγ-secreting cells after overnight stimulation with the indicated peptide pools. **c**, PBMCs of 15 individuals who recovered from SARS were stimulated in parallel with peptide pools covering the N proteins of SARS-CoV and SARS-CoV-2, and the frequency of IFNγ-producing cells is shown. **d**, The composition of the SARS-CoV-2 response in each individual who recovered from SARS ($n = 23$) is shown as a percentage of the total detected response. N-1, light blue; N-2, dark blue; NSP7, orange; NSP13-1, light red; NSP13-2, red; NSP13-3, dark red. **e**, PBMCs of 8 individuals who recovered from SARS were stimulated with all peptides covering N, NSP7 and NSP13 of SARS-CoV-2 to detect cross-reactive responses. The numbers of cells that are reactive to the different peptide pools directly ex vivo and after in vitro expansion are shown.

CD4 or CD8 T cells (Table 1 and Fig. 2b). Table 1 summarizes the different T cell epitopes that were defined by both ELISpot and ICS for seven individuals who recovered from COVID-19. Notably, we observed that COVID-19-convalescent individuals developed T cells that were specific to regions that were also targeted by T cells from individuals who recovered from SARS. For example, the region of amino acids 101–120 of the N protein, which is a previously described CD4 T cell epitope in SARS-CoV-exposed individuals[11,22], also stimulated CD4 T cells in two COVID-19-convalescent individuals. Similarly, the region of amino acids 321–340 of the N protein contained epitopes that triggered CD4 and CD8 T cells in patients who recovered from either COVID-19 or from SARS[22]. The finding that patients who recovered from COVID-19 and SARS can mount T cell responses against shared viral determinants suggests that previous SARS-CoV infection can induce T cells that are able to cross-react against SARS-CoV-2.

SARS-CoV-2-specific T cells in patients with SARS

For the management of the current pandemic and for vaccine development against SARS-CoV-2, it is important to understand whether acquired immunity will be long-lasting. We have previously demonstrated that patients who recovered from SARS have T cells that are specific to epitopes within different SARS-CoV proteins that persist for 11 years after infection[11]. Here, we collected PBMCs 17 years after SARS-CoV infection and tested whether they still contained cells that were reactive against SARS-CoV and whether these had cross-reactive potential against SARS-CoV-2 peptides. PBMCs from individuals who had resolved a SARS-CoV infection ($n = 15$) were stimulated directly ex vivo with peptide pools that covered the N protein of SARS-CoV (N-1 and N-2), NSP7 and NSP13 (Fig. 3a). This revealed that 17 years after infection, IFNγ responses to SARS-CoV peptides were still present and were almost exclusively focused on the N protein rather than the NSP peptide pools (Fig. 3b). Subsequently, we tested whether the N peptides of SARS-CoV-2 (amino acid identity, 94%) induced IFNγ responses in PBMCs from individuals who resolved a SARS-CoV infection. Indeed, PBMCs from all 23 individuals tested reacted to N peptides from SARS-CoV-2 (Fig. 3c, d). To test whether these low-frequency responses in individuals who had recovered from SARS could expand after encountering the N protein of SARS-CoV-2, the quantity of IFNγ-producing cells that responded to the N, NSP7 and NSP13 proteins of SARS-CoV-2 was analysed after 10 days of cell culture in the presence of the relevant peptides. Seven out of eight individuals tested showed clear, robust expansion of N-reactive cells (Fig. 3e) and ICS confirmed that individuals who recovered from SARS had SARS-CoV N-reactive CD4 and CD8 memory T cells[11] (Extended Data Fig. 6). In contrast to the response to the N peptides, we could not detect any cells that reacted to the peptide pools that covered NSP13 and only cells from one out of eight individuals reacted to NSP7 (Fig. 3e).

Thus, SARS-CoV-2 N-specific T cells are part of the T cell repertoire of individuals with a history of SARS-CoV infection and these T cells are able to robustly expand after encountering N peptides of SARS-CoV-2. These findings demonstrate that virus-specific T cells induced by infection with betacoronaviruses are long-lasting, supporting the notion that patients with COVID-19 will develop long-term T cell immunity. Our findings also raise the possibility that long-lasting T cells generated

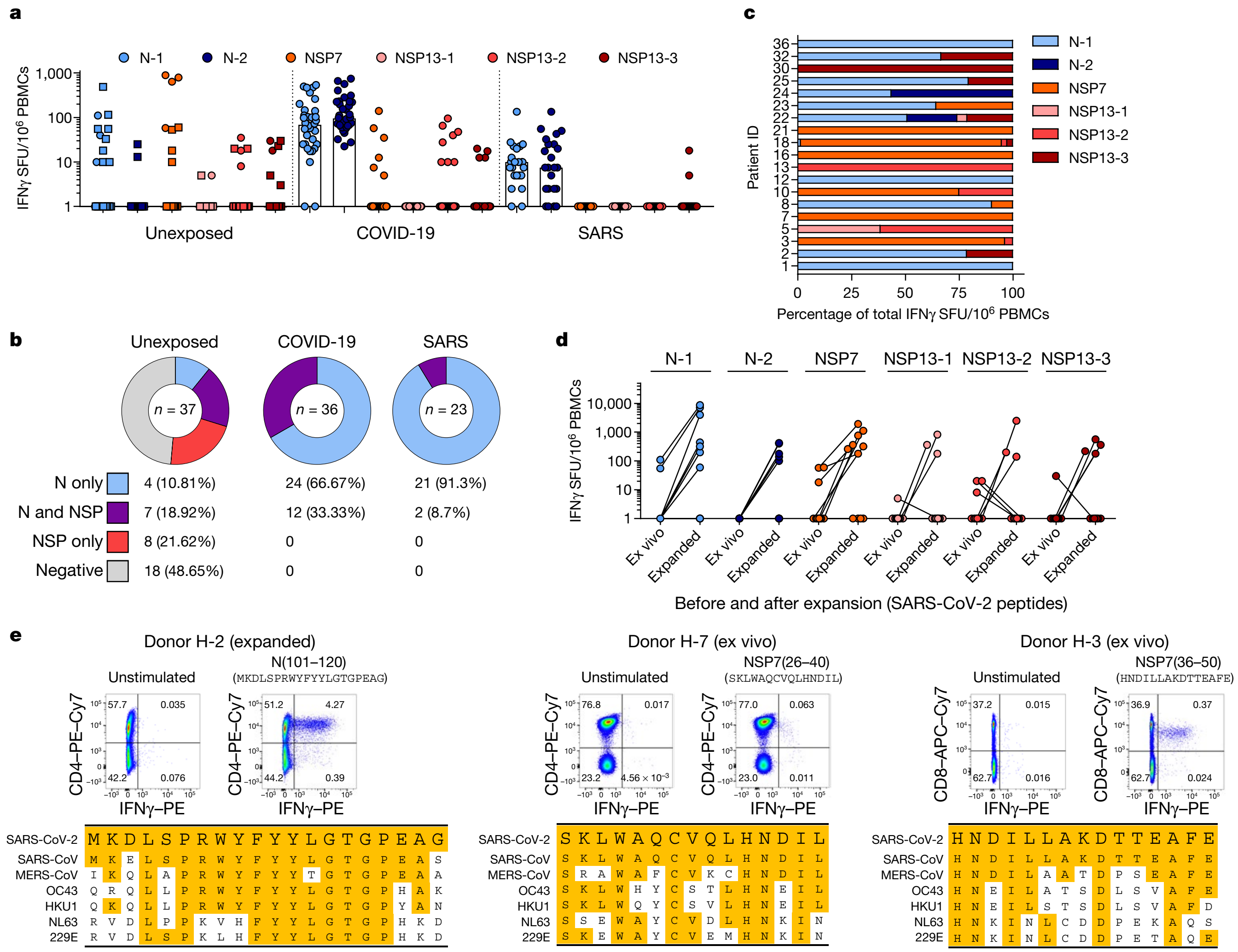


Fig. 4 | Immunodominance of SARS-CoV-2 responses in patients who recovered from COVID-19 and SARS, and in unexposed individuals. **a**, PBMCs of individuals who were not exposed to SARS-CoV and SARS-CoV-2 (*n* = 37), recovered from SARS (*n* = 23) or COVID-19 (*n* = 36) were stimulated with peptide pools covering N (N-1, N-2), NSP7 and NSP13 (NSP13-1, NSP13-2, NSP13-3) of SARS-CoV-2 and analysed by ELISpot. The frequency of peptide-reactive cells is shown for each donor (dots or squares) and the bars represent the median frequency. Squares denote PBMC samples collected before July 2019. **b**, The percentage of individuals with N-specific, NSP7 and NSP13-specific responses, or N-, NSP7- and NSP13-specific responses in cohort. **c**, The composition of the SARS-CoV-2 response in each responding unexposed donor (*n* = 19) is shown as a percentage of the total detected response. N-1, light blue; N-2, dark blue; NSP7, orange; NSP13-1, light red; NSP13-2, red; NSP13-3, dark red. **d**, Frequency of SARS-CoV-2-reactive cells in 11 unexposed donors to the indicated peptide pools directly ex vivo and after a 10-day expansion. **e**, A peptide pool matrix strategy was used for three individuals who were not exposed to SARS-CoV and SARS-CoV-2. The identified T cell epitopes were confirmed by ICS, and the sequences were aligned to the corresponding sequence of all coronaviruses known to infect humans.

after infection with related viruses may be able to protect against, or modify the pathology caused by, infection with SARS-CoV-2.

SARS-CoV-2-specific T cells in unexposed donors

To explore this possibility, we tested N-, NSP7- and NSP13-peptide-reactive IFNγ responses in 37 donors who were not exposed to SARS-CoV and SARS-CoV-2. Donors were either sampled before July 2019 (*n* = 26) or were serologically negative for both SARS-CoV-2 neutralizing antibodies and SARS-CoV-2 N antibodies[23] (*n* = 11). Different coronaviruses known to cause common colds in humans such as OC43, HKU1, NL63 and 229E present different degrees of amino acid homology with SARS-CoV-2 (Extended Data Fig. 1 and 2) and recent data have shown the presence of SARS-CoV-2 cross-reactive CD4 T cells (mainly specific to the spike protein) in donors who were not exposed to SARS-CoV-2[14]. Notably, we detected SARS-CoV-2-specific IFNγ responses in 19 out of 37 unexposed donors (Fig. 4a, b). The cumulative proportion of all studied individuals who responded to peptides covering the N protein and the ORF1-encoded NSP7 and NSP13 proteins is shown in Fig. 4b. Unexposed donors showed a distinct pattern of reactivity; whereas individuals who recovered from COVID-19 and SARS reacted preferentially to N peptide pools (66% of individuals who recovered from COVID-19 and 91% of individuals who recovered from SARS responded to only the N peptide pools), the unexposed group showed a mixed response to the N protein or to NSP7 and NSP13 (Fig. 4a–c). In addition, whereas NSP peptides stimulated a dominant response in only 1 out of 59 individuals who had resolved COVID-19 or SARS, these peptides triggered dominant reactivity in 9 out of 19 unexposed donors with SARS-CoV-2-reactive cells (Fig. 4c and Extended Data Fig. 7). These SARS-CoV-2-reactive cells from unexposed donors had the capacity to expand after stimulation with SARS-CoV-2-specific peptides (Fig. 4d). We next delineated the SARS-CoV-2-specific response detected in unexposed donors in more detail. Characterization of the N-specific response in one donor (H-2) identified CD4 T cells that were reactive to an epitope within the region of amino acids 101–120 of the N protein. This epitope was also detected in patients who recovered from COVID-19 and SARS[8,22] (Fig. 2b). This region has a high degree of homology to the sequences of the N protein of MERS-CoV, OC43 and HKU1 (Fig. 4e). In the same donor, we analysed PBMCs collected at multiple time points, demonstrating the persistence of the response to the 101–120 amino acid region of the N protein over 1 year (Extended Data Fig. 8a). In three other donors who were not exposed to SARS-CoV or SARS-CoV-2, we identified CD4 T cells specific to the region of amino acids 26–40 of NSP7 (SKLWAQCVQLHNDIL; donor H-7) and CD8 T cells specific to an epitope comprising the region of amino acids 36–50 of NSP7 (HNDILLAKDTTEAFE; H-3, H-21; Fig. 4e, Extended Data Fig. 8b).

These latter two T cell specificities were of particular interest as the homology between the two protein regions of SARS-CoV, SARS-CoV-2 and other common cold coronaviruses (OC43, HKU1 NL63 and 229E) was minimal (Fig. 4e), especially for the CD8 T cell epitope. Indeed, the low-homology peptides that covered the sequences of the common cold coronaviruses failed to stimulate PBMCs from individuals with T cells responsive to amino acids 36–50 of NSP7 (Extended Data Fig. 8c). Even though we cannot exclude that some SARS-CoV-2-reactive T cells might be naive or induced by completely unrelated pathogens[5], this finding suggests that unknown coronaviruses, possibly of animal origin, might induce cross-reactive SARS-CoV-2 T cells in the general population.

We further characterized the NSP7-specific CD4 and CD8 T cells that were present in the three unexposed individuals. The reactive T cells expanded efficiently in vitro and mainly produced either both IFNγ and TNF (CD8 T cells) or only IFNγ (CD4 T cells) (Extended Data Fig. 9a). We also determined that the CD8 T cells that were specific to amino acids 36–50 of NSP7 were HLA-B35-restricted and had an effector memory/terminal differentiated phenotype ($CCR7^{-}CD45RA^{+/-}$) (Extended Data Fig. 9b, c).

Conclusions

It is unclear why NSP7- and NSP13-specific T cells are detected and often dominant in unexposed donors, while representing a minor population in individuals who have recovered from SARS or COVID-19. It is, however, consistent with the findings of a previous study[11], in which ORF1-specific T cells were preferentially detected in some donors who were not exposed to SARS-CoV-2 whereas T cells from individuals who had recovered from COVID-19 preferentially recognized structural proteins. Induction of virus-specific T cells in individuals who were exposed but uninfected has been demonstrated in other viral infections[24–26]. Theoretically, individuals exposed to coronaviruses might just prime ORF1-specific T cells, as the ORF1-encoded proteins are produced first in coronavirus-infected cells and are necessary for the formation of the viral replicase–transcriptase complex that is essential for the subsequent transcription of the viral genome, which then leads to the expression of various RNA species[18]. Therefore, ORF1-specific T cells could hypothetically abort viral production by lysing SARS-CoV-2-infected cells before the formation of mature virions. By contrast, in patients with COVID-19 and SARS, the N protein—which is abundantly produced in cells that secrete mature virions[17]—would be expected to preferentially boost N-specific T cells.

Notably, the ORF1 region contains domains that are highly conserved among many different coronaviruses[9]. The distribution of these viruses in different animal species might result in periodic human contact that induces ORF1-specific T cells with cross-reactive abilities against SARS-CoV-2. Understanding the distribution, frequency and protective capacity of pre-existing structural or non-structural protein-associated SARS-CoV-2 cross-reactive T cells could be important for the explanation of some of the differences in infection rates or pathology observed during this pandemic. T cells that are specific to viral proteins are protective in animal models of airway infections[27,28], but the possible effects of pre-existing N- and/or ORF1-specific T cells on the differential modulation of SARS-CoV-2 infection will have to be carefully evaluated.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2550-z.

1. Welsh, R. M. & Selin, L. K. No one is naive: the significance of heterologous T-cell immunity. *Nat. Rev. Immunol*. **2**, 417–426 (2002).
2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
3. Raoult, D., Zumla, A., Locatelli, F., Ippolito, G. & Kroemer, G. Coronavirus infections: epidemiological, clinical and immunological features and hypotheses. *Cell Stress* **4**, 66–75 (2020).
4. Lim, M. Q. et al. Cross-Reactivity and anti-viral function of dengue capsid and NS3-specific memory T cells toward Zika virus. *Front. Immunol*. **9**, 2225 (2018).
5. Su, L. F., Kidd, B. A., Han, A., Kotzin, J. J. & Davis, M. M. Virus-specific $CD4^+$ memory-phenotype T cells are abundant in unexposed adults. *Immunity* **38**, 373–383 (2013).
6. Wen, J. et al. $CD4^+$ T cells cross-reactive with dengue and Zika viruses protect against Zika virus infection. *Cell Rep*. **31**, 107566 (2020).
7. Urbani, S. et al. Heterologous T cell immunity in severe hepatitis C virus infection. *J. Exp. Med*. **201**, 675–680 (2005).
8. Nickbakhsh, S. et al. Epidemiology of seasonal coronaviruses: establishing the context for the emergence of coronavirus disease 2019. *J. Infect. Dis*. **222**, 17–25 (2020).
9. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol*. **17**, 181–192 (2019).
10. Cao, W.-C., Liu, W., Zhang, P.-H., Zhang, F. & Richardus, J. H. Disappearance of antibodies to SARS-associated coronavirus after recovery. *N. Engl. J. Med*. **357**, 1162–1163 (2007).
11. Ng, O.-W. et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **34**, 2008–2014 (2016).
12. Wu, A. et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **27**, 325–328 (2020).
13. Ni, L. et al. Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity* **52**, 971–977 (2020).
14. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2020).
15. Li, C. K.-F. et al. T cell responses to whole SARS coronavirus in humans. *J. Immunol*. **181**, 5490–5500 (2008).
16. Zhao, J. et al. Recovery from the Middle East respiratory syndrome is associated with antibody and T-cell responses. *Sci. Immunol*. **2**, eaan5393 (2017).
17. Irigoyen, N. et al. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog*. **12**, e1005473 (2016).
18. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol*. **14**, 523–534 (2016).
19. Knoops, K. et al. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol*. **6**, e226 (2008).
20. Rivino, L. et al. Defining $CD8^+$ T cell determinants during human viral infection in populations of Asian ethnicity. *J. Immunol*. **191**, 4010–4019 (2013).
21. Tan, A. T. et al. Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. *J. Virol*. **82**, 10986–10997 (2008).
22. Oh, H. L. J. et al. Engineering T cells specific for a dominant severe acute respiratory syndrome coronavirus CD8 T cell epitope. *J. Virol*. **85**, 10464–10471 (2011).
23. Yong, S. E. F. et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. *Lancet Infect. Dis*. **20**, 809–815 (2020).
24. Rowland-Jones, S. L. et al. HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant. *Lancet* **341**, 860–861 (1993).
25. Park, S.-H. et al. Subinfectious hepatitis C virus exposures suppress T cell responses against subsequent acute infection. *Nat. Med*. **19**, 1638–1642 (2013).
26. Werner, J. M., Abdalla, A., Gara, N., Ghany, M. G. & Rehermann, B. The hepatitis B vaccine protects re-exposed health care workers, but does not provide sterilizing immunity. *Gastroenterology* **145**, 1026–1034 (2013).
27. Zhao, J. et al. Airway memory $CD4^+$ T cells mediate protective immunity against emerging respiratory coronaviruses. *Immunity* **44**, 1379–1391 (2016).
28. McKinstry, K. K. et al. Memory $CD4^+$ T cells protect against influenza through multiple synergizing mechanisms. *J. Clin. Invest*. **122**, 2847–2856 (2012).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ethics statement

All donors provided written consent. The study was conducted in accordance with the Declaration of Helsinki and approved by the NUS Institutional Review Board (H-20-006) and the SingHealth Centralised Institutional Review Board (reference CIRB/F/2018/2387).

Human samples

Donors were recruited based on their clinical history of SARS-CoV or SARS-CoV-2 infection. Blood samples of patients who recovered from COVID-19 ($n = 36$) were obtained 2–28 days after PCR negativity and of patients who recovered from SARS ($n = 23$) 17 years after infection. Samples from healthy donors were either collected before June 2019 for studies of T cell function in viral diseases ($n = 26$), or in March–April 2020. All healthy donor samples tested negative for RBD-neutralizing antibodies and negative in an ELISA for N IgG ($n = 11$)[19].

PBMC isolation

PBMCs were isolated by density-gradient centrifugation using Ficoll–Paque. Isolated PBMCs were either studied directly or cryopreserved and stored in liquid nitrogen until use in the assays.

Peptide pools

We synthesized 15-mer peptides that overlapped by 10 amino acids and spanned the entire protein sequence of the N, NSP7 and NSP13 proteins of SARS-CoV-2, as well as the N protein of SARS-CoV (GL Biochem Shanghai; see Supplementary Tables 1, 2). To stimulate PBMCs, the peptides were divided into 5 pools of about 40 peptides covering N (N-1, N-2) and NSP13 (NSP13-1, NSP13-2, NSP13-3) and one pool of 15 peptides covering NSP7. For single-peptide identification, peptides were organized in a matrix of 12 numeric and 7 alphabetical pools for N, and 4 numeric and 4 alphabetical pools for NSP7.

ELISpot assay

ELISpot plates (Millipore) were coated with human IFNγ antibody (1-D1K, Mabtech; 5 μg/ml) overnight at 4 °C. Then, 400,000 PBMCs were seeded per well and stimulated for 18 h with pools of SARS-CoV or SARS-CoV-2 peptides (2 μg/ml). For stimulation with peptide matrix pools or single peptides, a concentration of 5 μg/ml was used. Subsequently, the plates were developed with human biotinylated IFNγ detection antibody (7-B6-1, Mabtech; 1:2,000), followed by incubation with streptavidin-AP (Mabtech) and KPL BCIP/NBT Phosphatase Substrate (SeraCare). Spot forming units (SFU) were quantified with ImmunoSpot. To quantify positive peptide-specific responses, 2× mean spots of the unstimulated wells were subtracted from the peptide-stimulated wells, and the results expressed as SFU/10^6 PBMCs. We excluded the results if negative control wells had >30 SFU/10^6 PBMCs or positive control wells (phorbol 12-myristate 13-acetate/ionomycin) were negative.

Flow cytometry

PBMCs or expanded T cell lines were stimulated for 5 h at 37 °C with or without SARS-CoV or SARS-CoV-2 peptide pools (2 μg/ml) in the presence of 10 μg/ml brefeldin A (Sigma-Aldrich). Cells were stained with the yellow LIVE/DEAD fixable dead cell stain kit (Invitrogen) and anti-CD3 (clone SK7; 3:50), anti-CD4 (clone SK3; 3:50) and anti-CD8 (clone SK1; 3:50) antibodies. For analysis of the T cell differentiation status, cells were additionally stained with anti-CCR7 (clone 150503; 1:10) and anti-CD45RA (clone HI100; 1:10) antibodies. Cells were subsequently fixed and permeabilized using the Cytofix/Cytoperm kit (BD Biosciences-Pharmingen) and stained with anti-IFNγ (clone 25723, R&D Systems; 1:25) and anti-TNF (clone MAb11; 1:25) antibodies and analysed on a BD-LSR II FACS Scan. Data were analysed by FlowJo (Tree Star). Antibodies were purchased from BD Biosciences-Pharmingen unless otherwise stated.

Expanded T cell lines

T cell lines were generated as follows: 20% of PBMCs were pulsed with 10 μg/ml of the overlapping SARS-CoV-2 peptides (all pools combined) or single peptides for 1 h at 37 °C, washed and cocultured with the remaining cells in AIM-V medium (Gibco; Thermo Fisher Scientific) supplemented with 2% AB human serum (Gibco; Thermo Fisher Scientific). T cell lines were cultured for 10 days in the presence of 20 U/ml of recombinant IL-2 (R&D Systems).

HLA-restriction assay

The HLA type of healthy donor H-3 was determined and different Epstein–Barr virus (EBV)-transformed B cells lines with one common allele each were selected for presentation of peptide NSP7(36–50) (see below). B cells were pulsed with 10 μg/ml of the peptide for 1 h at 37 °C, washed three times and cocultured with the expanded T cell line at a ratio of 1:1 in the presence of 10 μg/ml brefeldin A (Sigma-Aldrich). Non-pulsed B cell lines served as a negative control for the detection of potential allogeneic responses and autologous peptide-pulsed cells served as a positive control. The HLA class I haplotype of the different B cell lines: CM780, A*24:02, A*33:03, B*58:01, B*55:02, Cw*07:02, Cw*03:02; WGP48, A*02:07, A*11:01, B*15:25, B*46:01, Cw*01:02, Cw*04:03; NP378, A*11:01, A*33:03, B*51:51, B*35:03, Cw*07:02, Cw*14:02; NgaBH, A*02:01, A*33:03, B*58:01, B*13:01, Cw*03:02.

Sequence alignment

Reference protein sequences for ORF1ab (accession numbers: QHD43415.1, NP_828849.2, YP_009047202.1, YP_009555238.1, YP_173236.1, YP_003766.2 and NP_073549.1) and the N protein (accession numbers: YP_009724397.2, AAP33707.1, YP_009047211.1, YP_009555245.1, YP_173242.1, YP_003771.1 and NP_073556.1) were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/). Sequences were aligned using the MUSCLE algorithm with default parameters and percentage identity was calculated in Geneious Prime 2020.1.2 (https://www.geneious.com). Alignment figures were made in Snapgene 5.1 (GSL Biotech).

Surrogate virus neutralization assay

A surrogate virus-neutralization test was used. Specifically, this test measures the quantity of anti-spike antibodies that block protein–protein interactions between the receptor-binding domain of the spike protein and the human ACE2 receptor using an ELISA-based assay[29].

Statistical analyses

All statistical analyses were performed in Prism (GraphPad Software); details are provided in the figure legends.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Reference protein sequences for ORF1ab (accession numbers: QHD43415.1, NP_828849.2, YP_009047202.1, YP_009555238.1, YP_173236.1, YP_003766.2 and NP_073549.1) and the N protein (accession numbers: YP_009724397.2, AAP33707.1, YP_009047211.1, YP_009555245.1, YP_173242.1, YP_003771.1 and NP_073556.1) were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/

protein/). All data are available in the Article or the Supplementary Information. Source data are provided with this paper.

29. Tan, C. W. et al. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-020-0631-z (2020).

Acknowledgements We thank M. K. Maini and S. Vasudevan for critical reading and editing of the manuscript. Grant support was provided by a Special NUHS COVID-19 Seed Grant Call, Project NUHSRO/2020/052/RO5+5/NUHS-COVID/6 (WBS R-571-000-077-733).

Author contributions N.L.B. and A.T.T. designed all experiments and analysed all of the data, prepared the figures and edited the paper; K.K., C.Y.L.T., M.H., A.C., M.L. and N.T. performed ELISpots and intracellular cytokine staining, and generated short-term T cell lines; M.H.Y.C. and M.L. performed viral sequence homology and analysed data; W.N.C. and L.-F.W. carried out antibody testing; M.I-C.C., E.E.O., S.K., P.A.T., J.G.-H.L. and Y.-J.T. selected and recruited patients and analysed clinical data; Y.-J.T. provided funding and designed the study; AB designed and coordinated the study, provided funding, analysed the data and wrote the paper.

Competing interests A.B. is a cofounder of Lion TCR, a biotechnology company that develops T cell receptors for the treatment of virus-related diseases and cancers. All other authors have no competing interests related to the study.

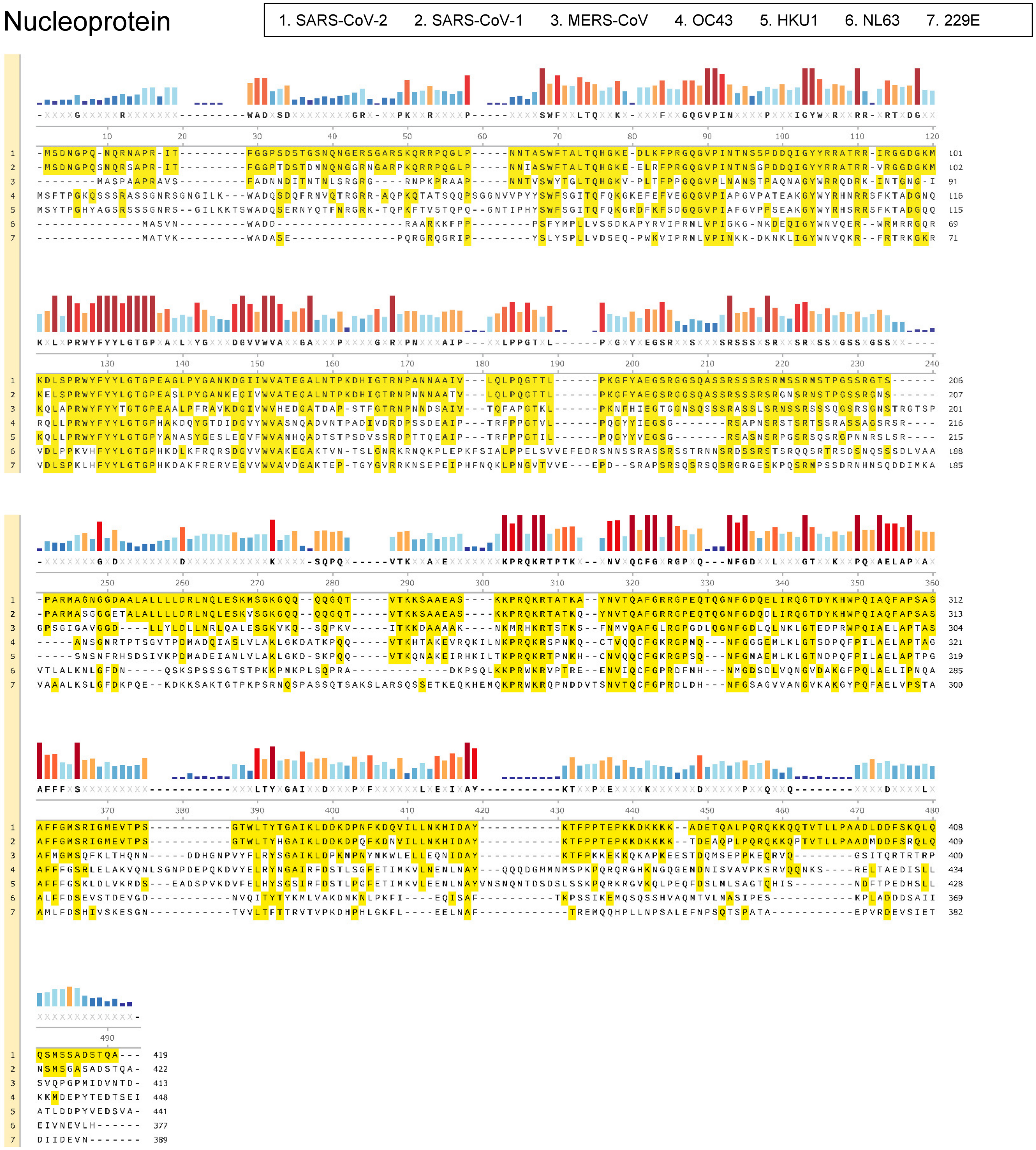
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2550-z.

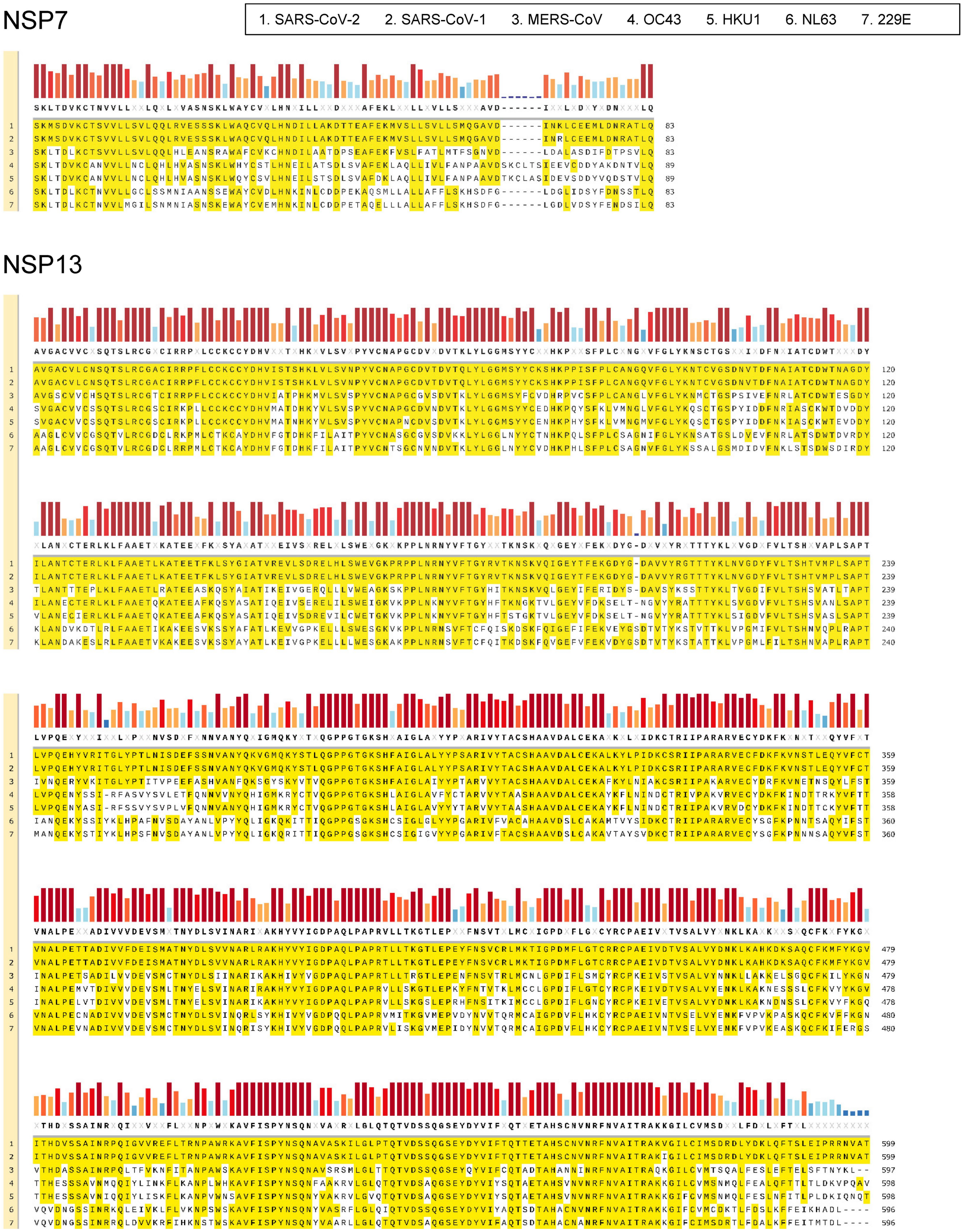
Correspondence and requests for materials should be addressed to A.B.

Peer review information *Nature* thanks Petter Brodin, Stanley Perlman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

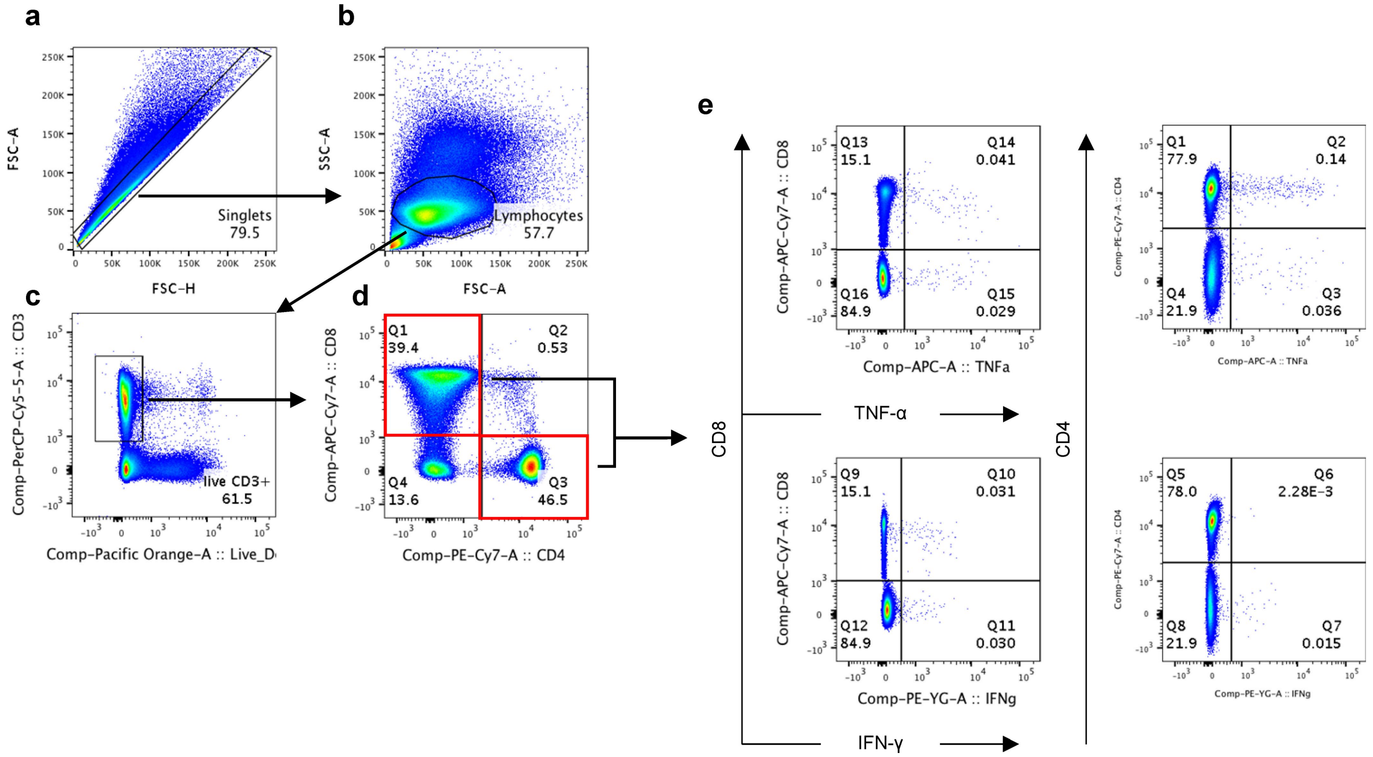
Reprints and permissions information is available at http://www.nature.com/reprints.



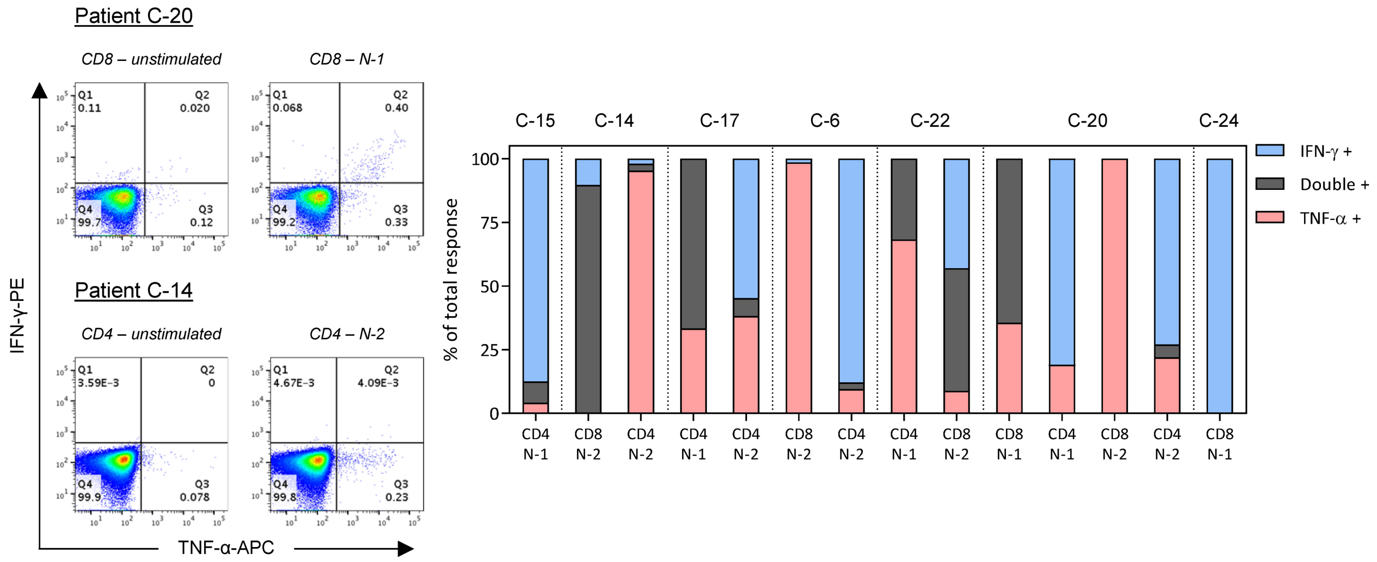
Extended Data Fig. 1 | Sequence alignment of the N protein from all types of human coronaviruses. Amino acid sequences for the N protein were downloaded from the NCBI database and aligned using the MUSCLE algorithm. Conserved residues are highlighted in yellow and the degree of conservation is indicated by the coloured bars above.



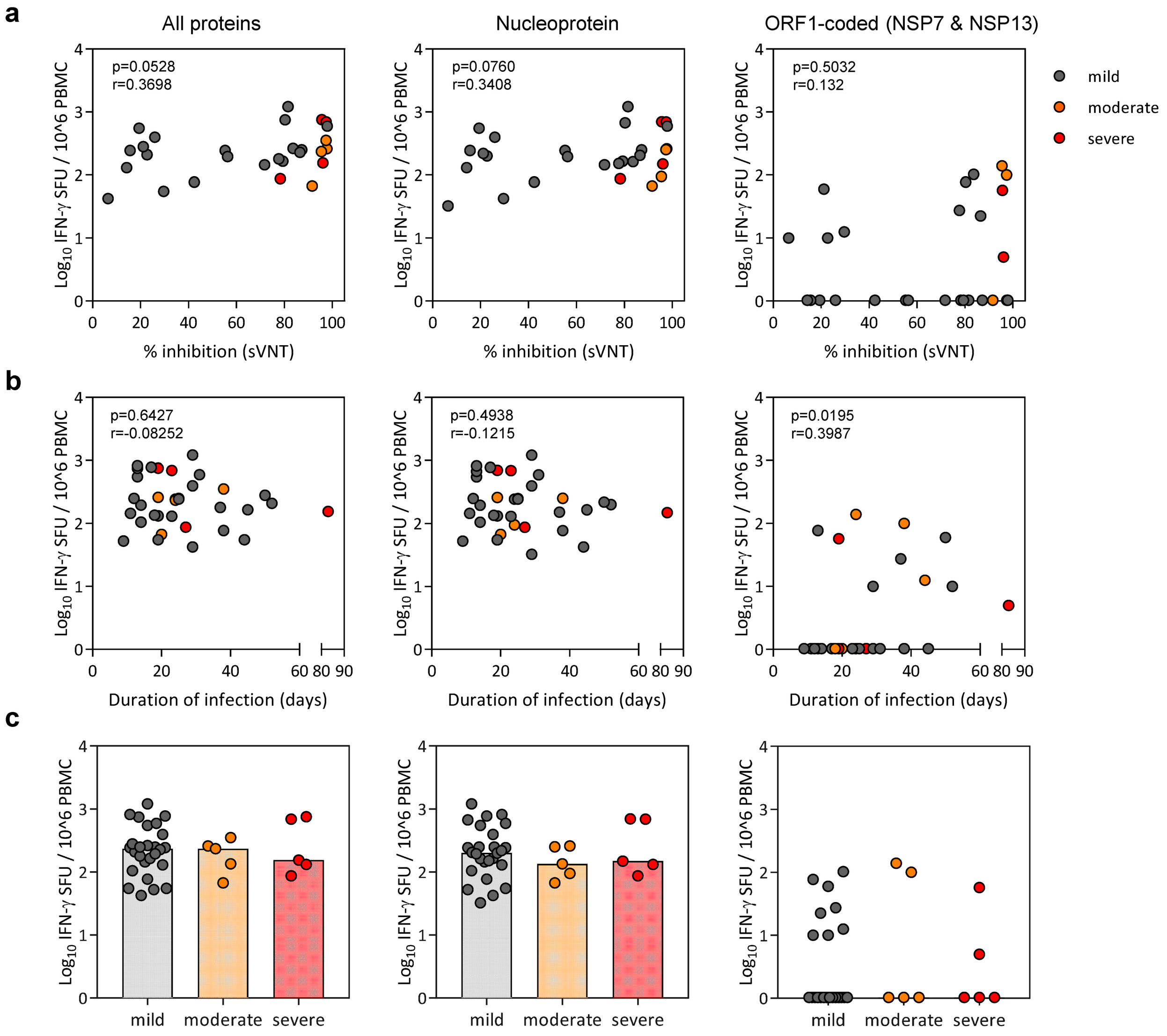
Extended Data Fig. 2 | Sequence alignment of the ORF1-encoded non-structural proteins NSP7 and NSP13 from all types of human coronaviruses. Protein sequences for ORF1ab were downloaded from the NCBI database and aligned using the MUSCLE algorithm. The alignment for NSP7 and NSP13 is shown.



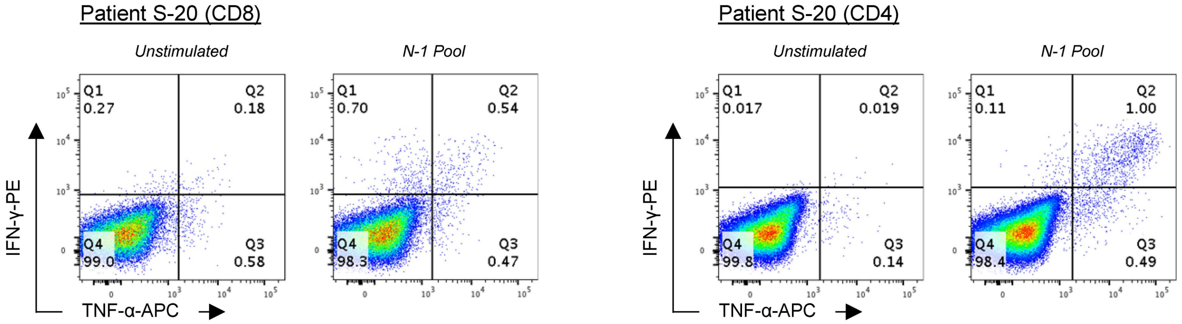
Extended Data Fig. 3 | Flow cytometry gating strategy. a, Forward scatter area (FSC-A) versus forward scatter height (FSC-H) density plot for doublet exclusion. **b**, Forward and side scatter (SSC-A) density plots to identify the lymphocyte population. **c**, Live T cells were gated based on CD3 expression and a live/dead discrimination dye. **d**, **e**, Only single expressing CD8 and CD4 T cells were Boolean gated (**d**) and used for IFNγ and/or TNF analysis (**e**).



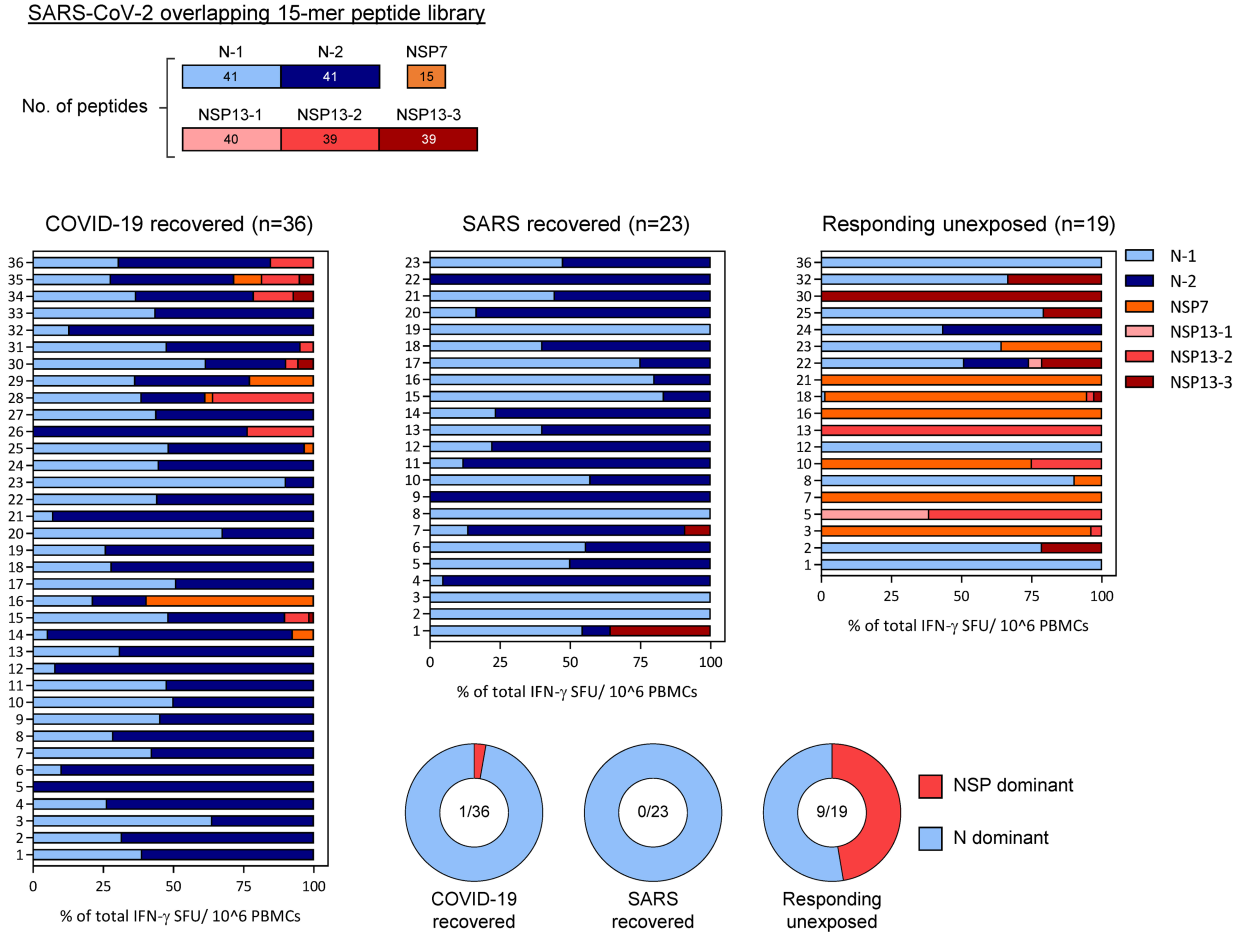
Extended Data Fig. 4 | IFNγ and TNF production profile of SARS-CoV-2-specific T cells of patients who recovered from COVID-19. PBMCs from patients recovered from COVID-19 ($n = 7$) were stimulated with the peptide pools covering N (NP-1, NP-2) for 5 h and analysed by intracellular cytokine staining for IFNγ and TNF. Dot plots show examples of patients with CD8 (top) or CD4 (bottom) T cells that produced IFNγ and/or TNF in response to stimulation with N-1 or N-2 peptide pools. The bars show the respective single and double cytokine producing T cells as a proportion of the total detected response after stimulation with the corresponding N peptide pools in each patient who recovered from COVID-19.



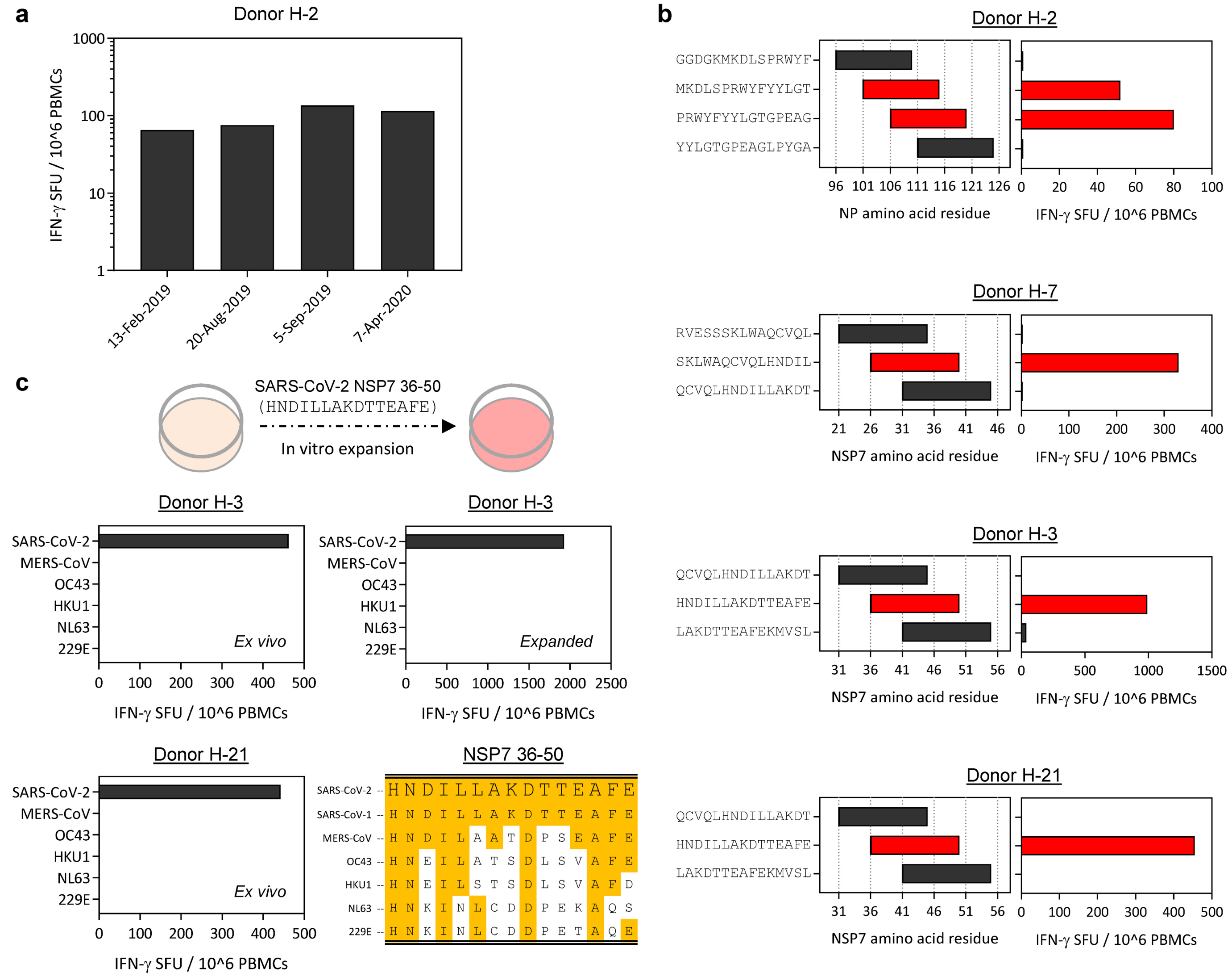
Extended Data Fig. 5 | Correlation analysis of SARS-CoV-2-specific IFNγ responses with the presence of virus-neutralizing antibodies, duration of infection and disease severity. a, **b**, The magnitude of SARS-CoV-2-specific responses, as quantified by IFNγ ELISpot, against all (N, NSP7 and NSP13) SARS-CoV-2 proteins tested (left), N (middle) or NSP7 and NSP13 (right) was correlated with the level of virus-neutralizing antibodies assayed using a surrogate virus neutralization assay (**a**; *n* = 28) and the duration of SARS-CoV-2 PCR positivity (**b**; *n* = 34). The respective *P* values (two-tailed) and correlation coefficients (Spearman correlation) are indicated. Patients who present with mild (grey), moderate (orange) or severe (red) disease are indicated. **c**, Magnitude of SARS-CoV-2-specific responses stratified by mild (*n* = 26), moderate (*n* = 5) and severe (*n* = 5) disease. The bars represent the median magnitude of the response. Mild disease, with or without chest radiograph changes, not requiring oxygen supplement. Moderate disease, oxygen supplement less than 50%. Severe disease, oxygen supplement 50% or more or high-flow oxygen or intubation.



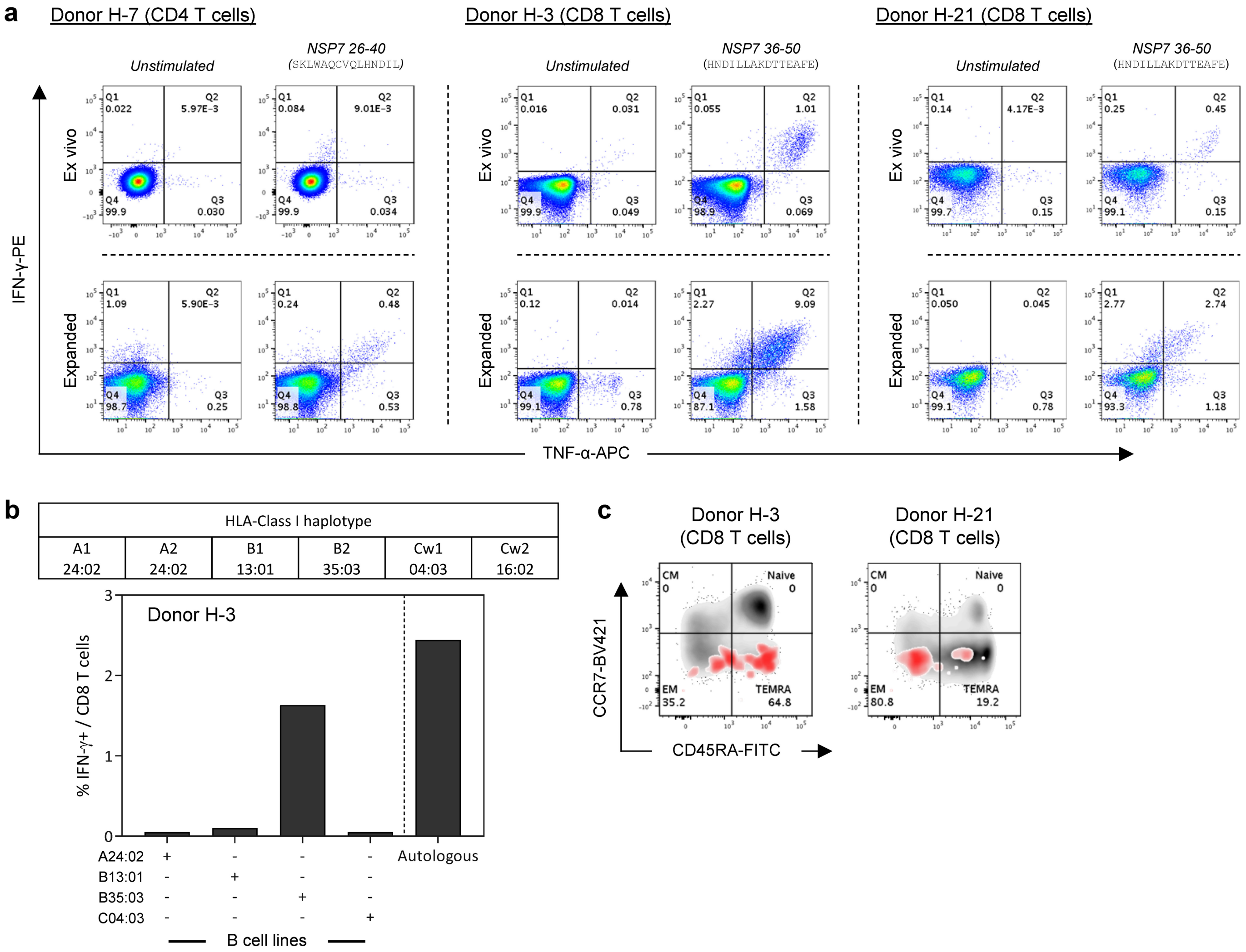
Extended Data Fig. 6 | Analysis of SARS-CoV N response. PBMCs of patient S-20 were expanded for 10 days and the frequency of T cells specific for the N-1 peptide pool were analysed by intracellular cytokine staining for IFNγ and TNF. Dot plots show CD8 and CD4 T cells that produced IFNγ and/or TNF in response to stimulation with the N-1 peptide pool.



Extended Data Fig. 7 | Dominance of SARS-CoV-2 N, NSP7 and NSP13 responses in donors who recovered from COVID-19 or SARS as well as in unexposed individuals. PBMCs from the respective individuals were stimulated with SARS-CoV-2 peptide pools as described in Fig. 1. The composition of the SARS-CoV-2 response is shown as a percentage of the total detected response in each group. N-1, light blue; N-2, dark blue; NSP7, orange; NSP13-1, light red; NSP13-2, red; NSP13-3, dark red. The proportion of individuals with NSP-dominant responses are illustrated in the pie charts.



Extended Data Fig. 8 | Identification of SARS-CoV-2 epitopes in donors who were not exposed to SARS-CoV and SARS-CoV-2. a, Longitudinal analysis of the SARS-CoV-2 N(101–120) response in individual H-2. PBMCs collected at the stated time points were stimulated with peptides spanning amino acids 101–120 of the N protein and assayed by IFNγ ELISpot. The frequencies of IFNγ SFU are shown. **b**, PBMCs were stimulated with the single peptides identified by the peptide matrix in parallel with the neighbouring peptides and assayed by IFNγ ELISpot. The amino acid residues are shown on the left; the frequency of IFNγ SFU on the right. Activating peptides are indicated in red and neighbouring peptides in black. **c**, PBMCs from individuals H-3 and H-21 were stimulated with the NSP7 peptide comprising amino acids 36–50 from SARS-CoV-2, MERS-CoV, OC43, HKU1, NL63 and 229E and analysed ex vivo by IFNγ ELISpot. A NSP7 (36–50) T cell line expanded from individual H-3 was also tested with the corresponding peptides of other coronaviruses by IFNγ ELISpot. Amino acid sequences of the various peptides are shown in the table. Conserved amino acids are highlighted in yellow.



Extended Data Fig. 9 | Characterization of SARS-CoV-2 NSP7-specific T cell responses in three individuals who were not exposed to SARS-CoV and SARS-CoV-2. a, Dot plots show the frequency of IFNγ- and/or TNF-producing CD8 or CD4 T cells specific to the SARS-CoV-2 peptides directly ex vivo and after a 10-day expansion in three unexposed donors. **b**, The HLA class I haplotype of individual H-3 is shown in the table. HLA restriction of the NSP7(36–50)-specific T cells from this individual was deduced by co-culturing the T cells with NSP7(36–50)-peptide-pulsed EBV-transformed B cell lines that share the indicated HLA class I molecule (+). Activation of the NSP7(36–50)-specific T cells by autologous cells was achieved by the direct addition of the peptide and used as the positive control. **c**, The memory phenotype of CD8 T cells specific for NSP7(36–50) in individuals H-3 and H-21 were analysed ex vivo and shown in the dot plots. The frequencies of naive, effector memory, central memory and terminally differentiated NSP7(36–50)-specific CD8 T cells (red) are shown and density plots were overlaid on the total CD8 T cells (grey).

Extended Data Table 1 | Donor characteristics

	COVID-19 recovered	SARS recovered	SARS-CoV-1/2 unexposed
Number	36	23	37
Median age in years	42	49	39
(range)	(27-78)	(21-67)	(28-63)
Gender			
Male	72% (26/36)	26% (6/23)	62% (23/37)
Female	28% (10/36)	74% (17/23)	38% (14/37)
Residence			
Singapore	100%	100%	100%
Ethnicity			
Chinese	38.9% (14/36)	43.5% (10/23)	62.2% (23/37)
Caucasian	27.8% (10/36)	0% (0/23)	16.2% (6/37)
Indian	25.0% (9/36)	21.7% (5/23)	8.1% (3/37)
Bangladeshi	5.6% (2/36)	0% (0/23)	0% (0/37)
Japanese	2.8% (1/36)	0% (0/23)	0% (0/37)
Malay	0% (0/36)	30.4% (7/23)	13.5% (5/37)
Ceylonese	0% (0/36)	4.3% (1/23)	0% (0/37)
*Disease Severity			
Mild	72.2% (26/36)	73.9% (17/23)	N/A
Moderate	13.9% (5/36)	13% (3/23)	N/A
Severe	13.9% (5/36)	13% (3/23)	N/A
Critical	0% (0/24)	0	N/A
Virological parameters			
SARS-CoV-1 PCR positive	N/A	100%	N/A
SARS-CoV-2 PCR positivity	100%	N/A	N/A
[23]SARS-CoV-2 NP Ig positivity	100%	100%	0%
[23]SARS-CoV-2 RBD Ig positivity	100%	0%	0%
Time since PCR negativity	2-28 days	17 years	N/A

*Disease severity is defined as follows. Mild, with or without chest radiograph changes; not requiring oxygen supplement. Moderate, oxygen supplement less than 50%. Severe, oxygen supplement 50% or more or high-flow oxygen or intubation.

nature research

Corresponding author(s): Antonio Bertoletti

Last updated by author(s): Jul 3, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Graphpad Prism 7; Flowjo Version 10.6.2; ImmunoSpot 7.0.26.0 Viral sequences were aligned using the MUSCLE algorithm (3.8.425) with default parameters and percentage identity was calculated in Geneious Prime 2020.1.2 (https://www.geneious.com). Alignment figures were made in Snapgene 5.1 (GSL Biotech).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coronavirus reference protein sequences for ORF1ab and Nucleocapsid Protein were downloaded from the NCBI database. All other data are included in this manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Aim of the study was to characterize SARS-CoV-2-specific T cells in patients who recovered from SARS 17 years ago. 23 of those individuals gave informed consent and were available to donate blood samples. Therefore similar numbers of COVID-19 convalescents and non-infected controls were selected.
Data exclusions	No data points were excluded.
Replication	We evaluated the SARS-CoV-2 specific T cell responses in 36 COVID-19 convalescents, in 23 SARS-recovered, and in 37 uninfected donors.
Randomization	No randomization was used in this study, since we are comparing 3 different well defined cohorts: COVID-19 convalescents, SARS recovered patients and SARS-CoV-1/2 non-exposed individuals.
Blinding	Blinding was not done for this study. The groups were defined by their infection history and studied by the investigators using standard protocols.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	ELISpot: IFN-γ coating antibody (clone: 1-D1K, MabTech, Cat. Nr. 3420-3-1000); biotinylated IFN-γ detection antibody (clone: 7-B6-1, MabTech, Cat. Nr: 3420-6-1000) Flow cytometry: anti-human CD3-PerCP-cy5.5 (BD Pharmingen, clone: SK7, Cat. Nr: 340949); anti-human CD4-PECy7 (BD Pharmingen, clone: SK3, Cat. Nr: 557852); anti-human CD8-APC-Cy7 (BD Pharmingen, clone: SK1, Cat. Nr: 557834); anti-human TNFa-APC (BD Pharmingen, clone: MAb11, Cat. Nr: 554514); anti-human IFNg-PE (R&D Systems, clone: 25273, Cat. Nr: IC285P); anti-human CCR7-BV421 (BD Pharmingen, clone: 150503, Cat. Nr: 562555); anti-human CD45RA-FITC (BD Pharmingen, clone: HI100, Cat. Nr: 555488)
Validation	All antibodies were obtained from commercial vendors and we based specificity on descriptions and information provided in corresponding Data Sheets available and provided by the Manufacturers.

Human research participants

Policy information about studies involving human research participants

Population characteristics	The characteristics of the human research participants are described in Extended Data Table 1 of the manuscript.
Recruitment	All donors were recruited based on the infection history. COVID-19 convalescents were previously PCR positive for SARS-CoV-2; SARS-recovered donors were tested PCR positive 17 years ago for SARS-CoV. Written informed consent was obtained from all subjects. All donors were recruited and resident in Singapore, were of mixed ethnicity and age.
Ethics oversight	Written informed consent was obtained from all subjects. The study was conducted in accordance with the Declaration of Helsinki and approved by the NUS institutional review board (H-20-006); SingHealth Centralised Institutional Review Board (reference CIRB/F/2018/2387)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	PBMC and T cell lines were prepared and stained according to standard protocols
Instrument	BD-LSR II FACS Scan
Software	Flowjo Version 10.6.2
Cell population abundance	N/A. No sorting was performed.
Gating strategy	Gating strategy: live cells (yellow LIVE/DEAD positive cells were excluded); singlets (SSC-H/SSC-A); Lymphocytes (FSC-A/SSCA); CD3+ (CD-3-PerPC-Cy5.5/CD8-APC-Cy7); CD4+ and CD8+ (CD4--PECy7/CD8-APC-Cy7); IFNg+ and TNFa+ gates were based on the unstimulated control sample.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Article

Longitudinal analyses reveal immunological misfiring in severe COVID-19

https://doi.org/10.1038/s41586-020-2588-y

Received: 23 June 2020

Accepted: 21 July 2020

Carolina Lucas[1,17], Patrick Wong[1,17], Jon Klein[1,17], Tiago B. R. Castro[2,17], Julio Silva[1], Maria Sundaram[3], Mallory K. Ellingson[3], Tianyang Mao[1], Ji Eun Oh[1], Benjamin Israelow[1,4], Takehiro Takahashi[1], Maria Tokuyama[1], Peiwen Lu[1], Arvind Venkataraman[1], Annsea Park[1], Subhasis Mohanty[4], Haowei Wang[4], Anne L. Wyllie[3], Chantal B. F. Vogels[3], Rebecca Earnest[3], Sarah Lapidus[3], Isabel M. Ott[3], Adam J. Moore[3], M. Catherine Muenker[3], John B. Fournier[4], Melissa Campbell[4], Camila D. Odio[4], Arnau Casanovas-Massana[3], Yale IMPACT Team*, Roy Herbst[5], Albert C. Shaw[4], Ruslan Medzhitov[1,6], Wade L. Schulz[7,8], Nathan D. Grubaugh[3], Charles Dela Cruz[9], Shelli Farhadian[4], Albert I. Ko[3,4], Saad B. Omer[3,4,10] & Akiko Iwasaki[1,6]✉

Recent studies have provided insights into the pathogenesis of coronavirus disease 2019 (COVID-19)[1–4]. However, the longitudinal immunological correlates of disease outcome remain unclear. Here we serially analysed immune responses in 113 patients with moderate or severe COVID-19. Immune profiling revealed an overall increase in innate cell lineages, with a concomitant reduction in T cell number. An early elevation in cytokine levels was associated with worse disease outcomes. Following an early increase in cytokines, patients with moderate COVID-19 displayed a progressive reduction in type 1 (antiviral) and type 3 (antifungal) responses. By contrast, patients with severe COVID-19 maintained these elevated responses throughout the course of the disease. Moreover, severe COVID-19 was accompanied by an increase in multiple type 2 (anti-helminths) effectors, including interleukin-5 (IL-5), IL-13, immunoglobulin E and eosinophils. Unsupervised clustering analysis identified four immune signatures, representing growth factors (A), type-2/3 cytokines (B), mixed type-1/2/3 cytokines (C), and chemokines (D) that correlated with three distinct disease trajectories. The immune profiles of patients who recovered from moderate COVID-19 were enriched in tissue reparative growth factor signature A, whereas the profiles of those with who developed severe disease had elevated levels of all four signatures. Thus, we have identified a maladapted immune response profile associated with severe COVID-19 and poor clinical outcome, as well as early immune signatures that correlate with divergent disease trajectories.

COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a highly infectious virus that exploits angiotensin-converting enzyme 2 (ACE2)[5,6] as a cell entry receptor. The clinical presentation of COVID-19 involves a broad range of symptoms and disease trajectories. Understanding the nature of the immune response that leads to recovery over severe disease is key to developing effective treatments for COVID-19. Coronaviruses, including Severe Acute Respiratory Syndrome (SARS-CoV) and Middle Eastern Respiratory Syndrome (MERS), typically induce strong inflammatory responses and associated lymphopenia[7,8]. Studies of patients with COVID-19 have reported increases in inflammatory monocytes and neutrophils, and a sharp decrease in lymphocytes[1–4], and an inflammatory milieu containing IL-1β, IL-6, and TNF (previously known as TNFα) in severe disease[1,2,4,9,10]. Despite these analyses, the dynamics of the immune response during the course of SARS-CoV-2 infection and its association with clinical trajectory remain unclear.

Immune responses against pathogens are divided roughly into three types[11–13]. Type 1 immunity, characterized by responses that depend on the transcription factor T-bet (also known as TBX21) and expression of interferon-γ (IFNγ), is generated against intracellular pathogens such as viruses. In type 1 immunity, pathogen clearance is mediated through effector cells including group 1 innate lymphocytes (ILC1), natural killer (NK) cells, cytotoxic T lymphocytes, and T helper 1 ($\mathrm{T_H}1$) cells. Type 2 immunity, which relies on the GATA3 transcription factor, mediates defence against helminths through effector molecules such as IL-4, IL-5, IL-13, and IgE that work to expel these pathogens through the concerted action of epithelial cells, mast cells, eosinophils, and basophils. Type 3 immunity, which is orchestrated by the RORγt-induced cytokines IL-17

---

[1] Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA. [2] Laboratory of Mucosal Immunology, The Rockefeller University, New York, NY, USA. [3] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA. [4] Department of Medicine, Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT, USA. [5] Yale University School of Medicine, Yale Cancer Center, and Smilow Cancer Hospital, New Haven, CT, USA. [6] Howard Hughes Medical Institute, Chevy Chase, MD, USA. [7] Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT, USA. [8] Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA. [9] Department of Medicine, Section of Pulmonary and Critical Care Medicine, Yale University School of Medicine, New Haven, CT, USA. [10] Yale Institute for Global Health, Yale University, New Haven, CT, USA. [17] These authors contributed equally: Carolina Lucas, Patrick Wong, Jon Klein, Tiago B. R. Castro. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: akiko.iwasaki@yale.edu

and IL-22 secreted by ILC3 and T_H17 cells, is mounted against fungi and extracellular bacteria to elicit neutrophil-dependent clearance. We have focused on the longitudinal analysis of these three types of immune response in patients with COVID-19 and identified correlations between distinct immune phenotypes and disease course.

Immunological features of COVID-19

One hundred and thirteen patients with COVID-19 who were admitted to Yale New Haven Hospital (YNHH) between 18 March 2020 and 27 May 2020 were recruited to the Yale IMPACT (Implementing Medical and Public Health Action Against Coronavirus CT) study. We assessed viral RNA load (quantified by quantitative PCR with reverse transcription (RT–qPCR) using nasopharyngeal swabs); levels of plasma cytokines and chemokines; and leukocyte profiles (by flow cytometry using freshly isolated peripheral blood mononuclear cells; PBMCs). We performed 253 collections and follow-up measurements on the patient cohort with a range of one to seven longitudinal time-points that occurred 3–51 days after the onset of symptoms. In parallel, we enrolled 108 volunteer healthcare workers (HCWs), whose samples served as healthy controls (SARS-CoV-2-negative by RT–qPCR and serology).

Basic demographic information stratified by disease severity is provided in Extended Data Table 1 and detailed in Supplementary Table 1. Patients who had been admitted to YNHH were stratified into moderate and severe disease groups on the basis of supplemental oxygen requirements and admission to the intensive care unit (ICU) (Fig. 1a). Among our cohort, patients who developed moderate or severe disease did not differ significantly with respect to age or sex. Body mass index (BMI) was generally higher among patients with severe disease, and extremes in BMI correlated with an increased relative risk (RR) of mortality (RR BMI ≥ 35: 1.62 (95% confidence interval (CI) 0.81–3.22)) (Extended Data Table 1, Extended Data Fig. 1a, b). Exposure to select therapeutic regimens of interest was assessed in patients with moderate or severe disease (Extended Data Fig. 1c.) Initial presenting symptoms demonstrated a preponderance of headache (54.55%), fever (64.47%), cough (74.03%), and dyspnoea (67.09%) with no significant difference in symptom presentation between patients with moderate disease and those who developed severe disease. Finally, mortality was significantly higher in patients who were admitted to the ICU than in those who were not (27.27% versus 3.75%; $P < 0.001$) (Extended Data Table 1).

We analysed PBMC and plasma samples from patients with moderate or severe COVID-19 and healthy HCW donors (Fig. 1a, gating strategy in Extended Data Fig. 9) by flow cytometry and ELISA to quantify leukocytes and soluble mediators, respectively. An unsupervised heat map constructed from the main innate and adaptive circulating immune cell types revealed marked changes in patients with COVID-19 compared to uninfected HCWs (Fig. 1b). As reported[1–4], patients with COVID-19 presented with marked reductions in the number and frequency of both $CD4^+$ and $CD8^+$ T cells, even after normalizing for age as a possible confounder (Extended Data Fig. 1d). Granulocytes, such as neutrophils and eosinophils, are normally excluded from the PBMC fraction following density gradient separation. However, low-density granulocytes are found in the PBMC layer of peripheral blood collected from patients with inflammatory diseases[14]. In patients with COVID-19, increases in monocytes, low-density neutrophils and eosinophils correlated with the severity of disease (Fig. 2c, Extended Data Fig. 2a, b). In addition, patients showed increased activation of T cells and a reduction in expression of the human leukocyte antigen DR isotype (HLA-DR) by circulating monocytes[1] (Extended Data Fig. 2c). A complete overview of PBMC subsets is presented in Extended Data Fig. 2.

To gain insights into key differences in cytokines, chemokines, and additional immune markers between patients with moderate and severe disease, we correlated the measurements of these soluble proteins across all sample collection time-points. (Fig. 1d). We observed a 'core COVID-19 signature' that was shared by both moderate and severe disease groups and was defined by the following inflammatory cytokines, which correlated positively with each other: IL-1α, IL-1β, IL-17A, IL-12 p70, and IFNα (Fig. 1d). In patients with severe disease, we observed an additional inflammatory cluster defined by thrombopoietin (TPO), IL-33, IL-16, IL-21, IL-23, IFNλ, eotaxin and eotaxin 3 (Fig. 1d). Most of the cytokines linked to cytokine release syndrome (CRS), such as IL-1α, IL-1β, IL-6, IL-10, IL-18 and TNF, showed increased positive associations in patients with severe disease (Fig. 1d–f, Extended Data Fig. 3). These data highlight broad inflammatory changes, involving concomitant release of type 1, type 2 and type 3 cytokines, in patients with severe COVID-19.

Longitudinal immune profiling of COVID-19

Our data presented above, as well as previous single-cell transcriptome and flow-cytometry-based studies[2,4,15–17], depicted overt innate and adaptive immune activation in patients with severe COVID-19. Longitudinal cytokine correlations, measured in terms of days from symptom onset (DfSO), indicated that major differences in immune phenotypes between moderate and severe disease were apparent after day 10 of infection (Fig. 2a). In the first 10 DfSO, patients with severe or moderate disease displayed similar correlation intensity and markers, including the overall core COVID-19 signature described above (Fig. 2a). After day 10 these markers declined steadily in patients with moderate disease. By contrast, patients with severe COVID-19 maintained elevated levels of these core signature makers. Notably, additional correlations between cytokines emerged in patients with severe disease following day 10 (Fig. 2a). These analyses strongly support the observation (Fig. 1) that TPO and IFNα associate strongly with IFNλ, IL-9, IL-18, IL-21, IL-23, and IL-33 (Fig. 2a). These observations indicate sharp differences in the expression of inflammatory markers along disease progression between patients who exhibit moderate versus severe symptoms of COVID-19.

Temporal analyses of PBMCs and soluble proteins in plasma, either by linear regression or grouped intervals, supported distinct courses in disease. IFNα levels were sustained at higher levels in patients with severe disease, but these declined in patients with moderate disease (Fig. 2b). Plasma levels of IFNλ increased during the first week of symptoms in patients with severe disease, and remained elevated in later phases (Fig. 2b). In addition, inflammasome-induced cytokines, such as IL-1β and IL-18, were also higher in patients with severe disease than in patients with moderate disease at most time-points analysed (Fig. 2c). IL-1 receptor antagonist (IL-1Ra), which is induced by IL-1R signalling as a negative feedback regulator[18], was also increased in patients with severe COVID-19 from day 10 of disease onset (Extended Data Fig. 4).

With respect to type 1 immunity, there was an increased number of monocytes at approximately 14 DfSO in patients with severe but not moderate COVID-19 (Fig. 2d). The innate cytokine IL-12, a key inducer of type-1 immunity[11,12], displayed a similar pattern to IFNγ—increasing over time in patients with severe disease but declining steadily in those with moderate disease (Fig. 2d). Intracellular cytokine staining showed that $CD4^+$ and $CD8^+$ T cells from patients with moderate disease secreted comparable amounts of IFNγ to those from patients with severe disease. Together with the severe T cell depletion seen in patients with severe disease (Fig. 1), our data suggest that secretion of IFNγ by non-T cells (ILC1, NK cells), or non-circulating T cells in tissues was the primary contributor to the enhanced levels observed in patients with severe disease (Extended Data Fig. 5).

Type-2 immune markers continued to increase over time in patients with severe COVID-19, as indicated by the strong correlations observed at late time points for these patients (Fig. 2a). Eosinophils and levels of eotaxin-2 increased in patients with severe disease and remained higher than in patients with moderate disease (Fig. 2e). Type 2 innate immune cytokines, including thymic stromal lymphopoietin (TSLP) and IL-33, did not show significant differences between patients with severe and moderate disease (Fig. 2e). Levels of hallmark type 2 cytokines, including IL-5 (associated with eosinophilia) and IL-13 (Fig. 2e), were higher

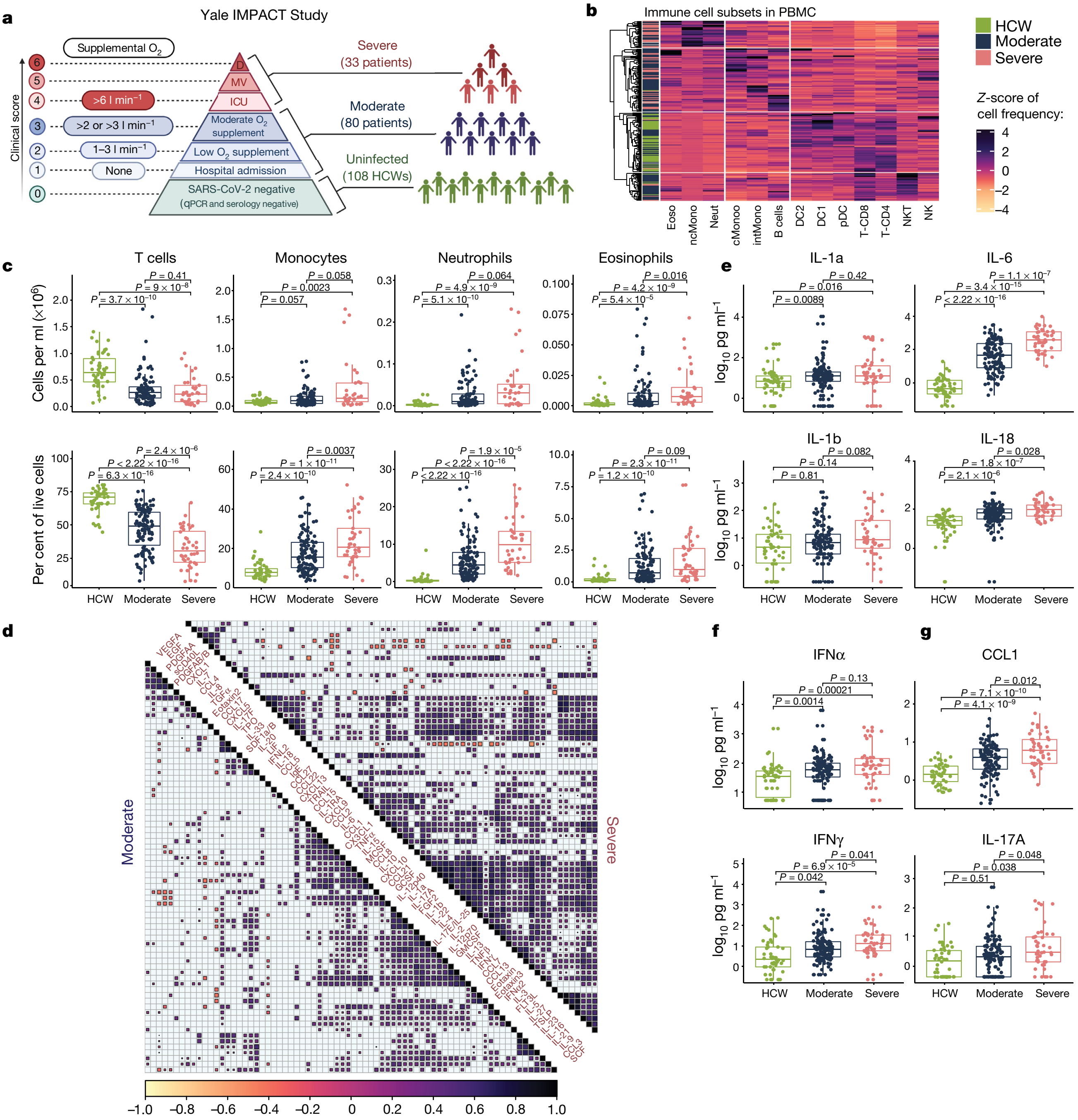


Fig. 1 | Overview of immunological features in patients with COVID-19. **a**, Overview of cohort, including healthy donors (HCWs) and patients with moderate or severe COVID-19. Ordinal scores assigned according to clinical severity scale as described in Methods. D, deceased; ICU, intensive care unit; MV, mechanical ventilation. **b**, Heat map comparison of the major immune cell populations within PBMCs in patients with moderate ($n = 121$) or severe ($n = 43$) COVID-19, or HCSs ($n = 43$). n values represent a separate time point per subject Subjects are arranged across rows, with each coloured unit indicating the relative distribution of an immune cell population normalized against the same population across all subjects. K-means clustering was used to arrange patients and measurements. Eoso, eosinophil; ncMono, non-classical monocyte; neut, neutrophil; cMono, classical monocyte; intMono, intermediate monocyte; DC2 and DC1, type 2 and 1 dendritic cells, respectively; pDC, plasmacytoid dendritic cell; T-CD8 and T-CD4, CD8$^+$ and CD4$^+$ T cells, respectively; NKT, natural killer T cell; NK, natural killer cell. **c**, Immune cell subsets plotted as a concentration of millions of cells per millilitre of blood or as a percentage of live single cells. Each dot represents a separate time point per subject (HCW, $n = 50$; moderate, $n = 117$; severe, $n = 40$). **d**, Correlation matrices across all time points of 71 cytokines from patient blood, comparing patients with moderate and severe disease. Only significant correlations (<0.05) are represented as dots. Pearson's correlation coefficients from comparisons of cytokine measurements within the same patients are visualized by colour intensity. **e**–**g**, Quantification of prominent inflammatory cytokines (**e**), interferons type I and II (**f**), and CCL1 and IL-17 (**g**) presented as $\log_{10}$-transformed concentrations. Each dot represents a separate time point per subject (HCW, $n = 50$; moderate, $n = 117$; severe, $n = 40$). Centre, median; box limits, first and third percentiles; whiskers, 1.5× interquartile range (IQR). Significance determined by two-sided, Wilcoxon rank-sum test.

in patients with severe disease than in those with moderate disease. By contrast, IL-4 levels were not significantly different. However, IL-4, similar to IL-5 and IL-13, showed an upward trend over the course of disease in patients with severe COVID-19 (Fig. 2e). The type 2 antibody isotype IgE was also higher in patients with severe diasease and continued to increase during the disease course (Fig. 2e).

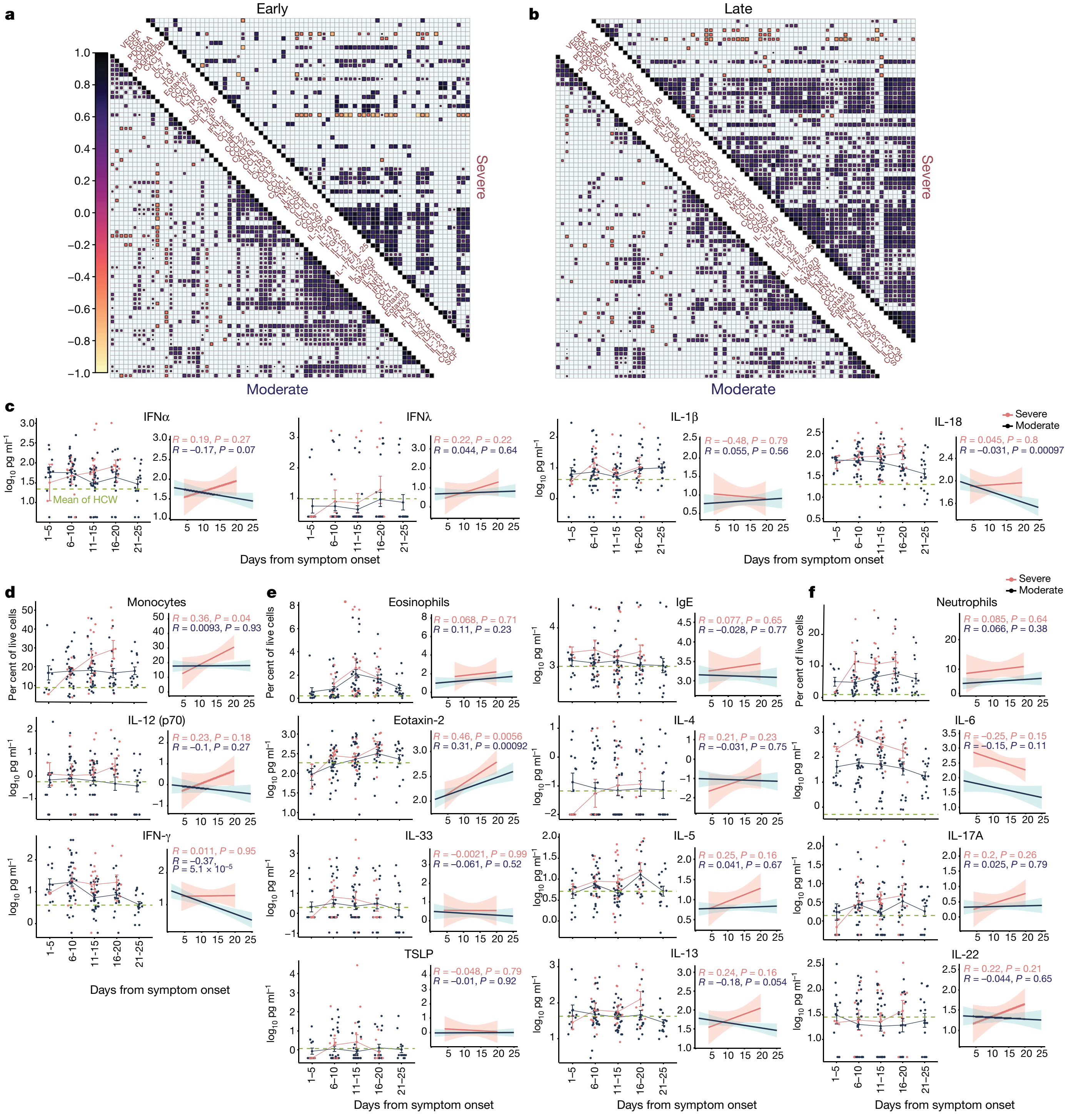


Fig. 2 | Longitudinal immune profiling of moderate and severe COVID-19 patients. a, Correlation matrices of 71 cytokines from patient blood comparing cytokine concentrations in patients with moderate or severe disease during the early phase (<10 DfSO) or late phase (>10 DfSO) of disease. Only significant correlations (<0.05) are represented as dots, and Pearson's correlation coefficient from comparisons of cytokine measurements within each patient is visualized by colour intensity. **b**, **c**, Anti-viral interferons (**b**) and inflammasome-related cytokines (**c**) plotted as $\log_{10}$ concentrations over time and grouped by disease severity. **d**–**f**, Cellular and cytokine measurements representative of type 1 (**d**), type 2 (**e**) and type 3 (**f**) immune responses reported over time in intervals of days (left) and continuously as linear regressions (right). Left, each dot represents a distinct patient and time point arranged in intervals of 5 days until 25 DfSO; dark blue, moderate disease ($n = 112$), pink, severe disease ($n = 40$). Dark blue or pink lines pass through the mean at each time interval; error bars denote the s.e.m. Dashed green line, mean from healthy HCWs. Right, regression lines are indicated by the dark blue (moderate) or red (severe) solid lines. Associated Pearson's correlation coefficients and linear regression significance are coloured accordingly; shading represents 95% CI.

IL-6, which is linked to CRS, was elevated in patients with severe disease[19]. Circulating neutrophils did not show a significant increase in our longitudinal analysis (Fig. 2f), although patients with severe disease showed hallmarks of type 3 responses, including increased plasma IL-17A and IL-22, as well as secretion of IL-17 by circulating CD4 T cells as assessed by intracellular cytokine staining (Fig. 2f, Extended Data Fig. 5). These data identify broad elevations of type 1, type 2 and type 3 signatures in severe cases of COVID-19, with differences in their

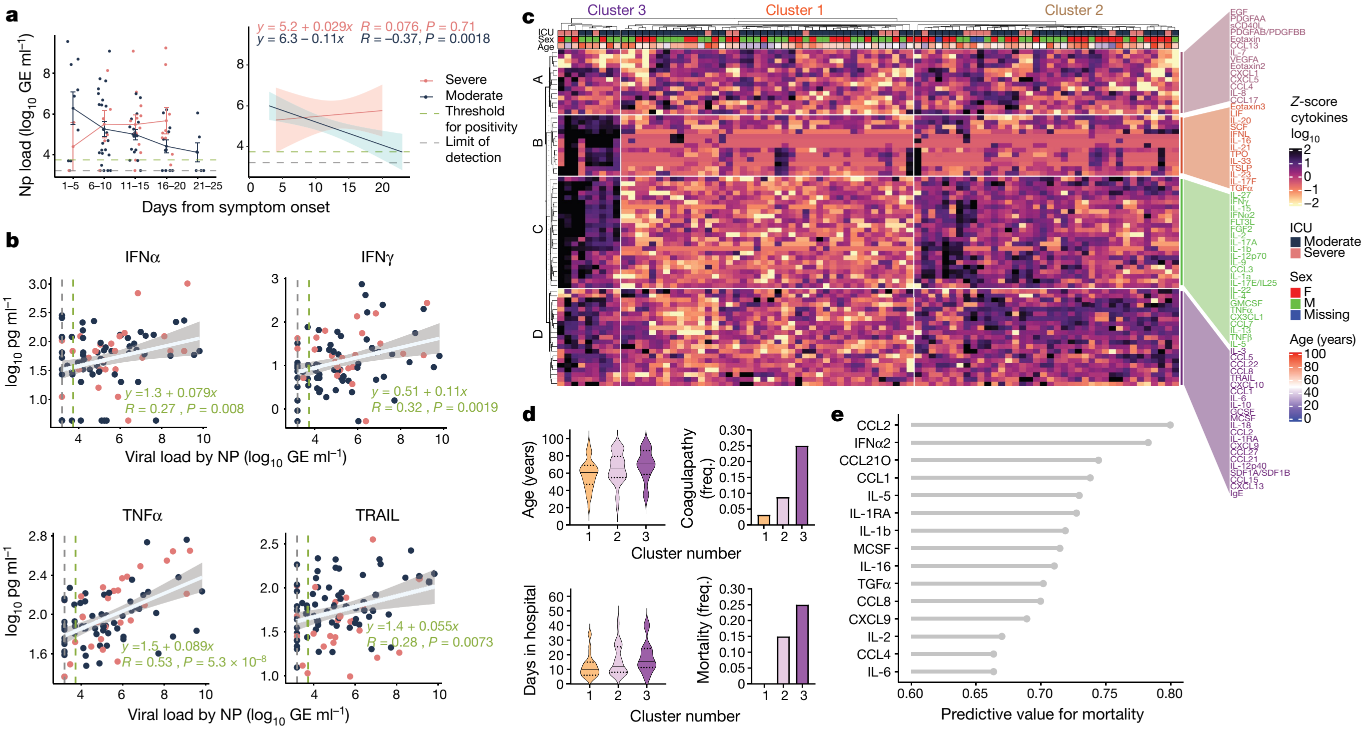


Fig. 3 | Early viral and cytokine profiles distinguish between moderate and severe disease outcomes. a, Viral loads measured by nasopharyngeal swabs are plotted as $\log_{10}$ of genome equivalents against time after symptom onset for patients with moderate disease ($n = 112$) or severe disease ($n = 39$). Left, each dot represents a distinct patient and time point arranged in intervals of 5 days until 25 DfSO. Dark blue or pink lines pass through the mean of each measurement; error bars denote s.e.m. Right, longitudinal data plotted over time continuously. Regression lines are shown as dark blue (moderate) or red (severe). Associated linear regression equations, Pearson's correlation coefficients, and significance are coloured accordingly. Green text is the regression analysis and correlation for all patients. Shading represents 95% CIs. Dashed green line denotes mean threshold for positivity. Dashed grey line indicates mean limit of detection. **b**, Correlation and linear regression of cytokines ($\log_{10}$ concentration) and viral load (by nasopharyngeal swab, $\log_{10}$ genome equivalents (GE)), regardless of disease severity ($n = 151$). Each dot represents a unique patient time point; dark blue, moderate disease; red, severe disease. White line indicates the regression line for all patients. The associated linear regression equation, Pearson's correlation coefficient, and significance are shown in green. Grey shading indicates 95% CIs. Dashed green line denotes mean threshold for positivity. Dashed grey line indicates mean limit of detection. **c**, Unbiased heat map comparisons of cytokines in PBMCs. Measurements were normalized across all patients. *K*-means clustering was used to determine clusters 1–3 (cluster 1, $n = 46$; cluster 2, $n = 50$; cluster 3, $n = 16$). **d**, Distribution of age and length of hospital stay (violin plots; solid lines, median; dotted lines, quartiles.) and frequency of coagulopathy and mortality (bar graphs) within each cluster. **e**, Top 20 cytokines by mutual information analysis to determine their importance for determining mortality. Significance of comparisons determined by two-sided, Wilcoxon rank-sum test.

kinetics and magnitudes between patients with severe and moderate disease.

Viral load correlates with elevated cytokines

We next measured viral load kinetics using serial nasopharyngeal swabs. Although there was no significant difference in viral RNA load between patients with moderate and severe disease at any specific time point analysed, patients with moderate disease showed a steady decline in viral load over the course of disease, whereas those with severe disease did not (Fig. 3a). Regardless of whether patients exhibited moderate or severe disease, viral load correlated significantly with the levels of IFNα, IFNγ, TNF and tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) (Fig. 3b). In addition, several chemokines responsible for monocyte recruitment correlated significantly with viral load only in patients with severe disease (Extended Data Fig. 6a, b). These data indicate that nasopharyngeal viral load correlates with plasma levels of interferons and cytokines.

Early cytokine profile marks disease outcomes

Next, we investigated whether specific early cytokine responses are associated with severe COVID-19. To this end, we conducted an unsupervised clustering analysis using baseline measurements collected before 12 DfSO (Fig. 3c). Three main clusters with correlation to distinct disease outcomes emerged. These were characterized by four distinct immune signatures. Signature A contained several stromal growth factors, including epidermal growth factor (EGF), platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF), that are mediators of wound healing and tissue repair[20], as well as IL-7, a key growth factor for lymphocytes. Signature B consisted of eotaxin 3, IL-33 and TSLP, along with IL-21, IL-23 and IL-17F, thus representing type 2 and type 3 immune effectors. Signature C comprised a mixture of all immunotypes, including type 1 (IFNγ, IL-12 p70, IL-15, IL-2 and TNF), type 2 (IL-4, IL-5 and IL-13), and type 3 cytokines (IL-1α, IL-1β, IL-17A, IL-17E and IL-22). Finally, signature D contained a number of chemokines involved in leukocyte trafficking, including CCL1, CCL2, CCL5, CCL8, CCL15, CCL21, CCL22, CCL27, CXCL9, CXCL10, CXCL13, and SDF1.

Cluster 1 primarily comprised patients with moderate disease who experienced low occurrences of coagulopathy, shorter lengths of hospital stay, and no mortality (Fig. 3c, d). The main characteristics in this cluster were low levels of inflammatory markers and similar or increased levels of parameters in signature A, which contains tissue reparative growth factors (Fig. 3c). Clusters 2 and 3 were characterized by a rise in inflammatory markers, and patients belonging to these clusters had a higher incidences of coagulopathy and mortality, which was more pronounced in cluster 3 (Fig. 3c, d). Patients in cluster 2 showed higher levels of markers in signatures C and D, which included IFNα, IL-1Ra and several hallmark type 1, type 2 and type 3 cytokines, than patients in cluster 1, but lower expression of markers in signatures B, C and D than those in

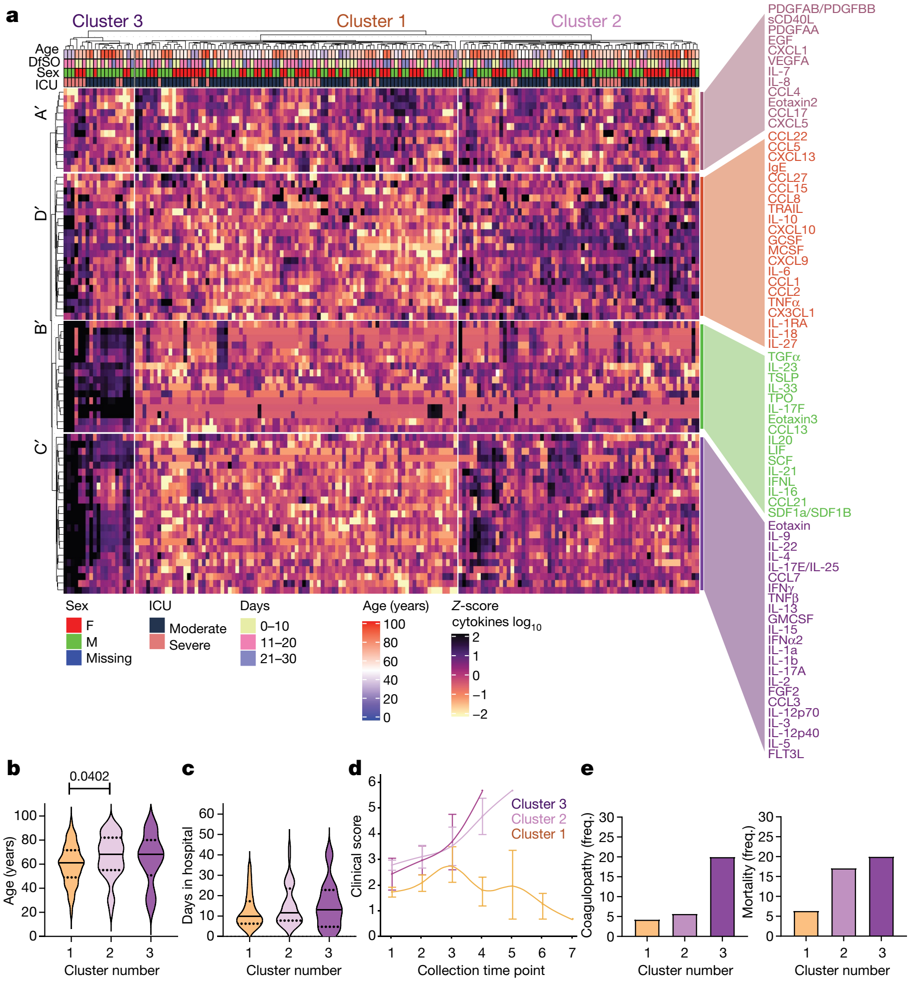


Fig. 4 | Immune correlates of COVID-19 outcomes. a, Unbiased heat map comparisons of cytokines in PBMCs measured at distinct time points in patients with COVID-19. Measurements were normalized across all patients. *K*-means clustering was used to determine clusters 1–3 (cluster 1, $n = 84$; cluster 2, $n = 66$; cluster 3, $n = 20$). **b**, **c**, Distribution of age (**b**) and length of hospital stay (violin plots) (**c**) of patients within each cluster. For statistical differences, adjusted *P* values calculated using one-way ANOVA with Tukey's correction for multiple comparisons are shown (age: $F_{(2,90)} = 3.115$; $P = 0.0492$). Solid lines, median; dotted lines, quartiles. **d**, Disease progression measured by clinical severity score for patients in each cluster. Data (mean ± s.e.m.) are ordered by the collection time points for each patient, with regular collection intervals of 3–4 days (Extended Data Fig. 7). **e**, Percentage of patients in each cluster with new-onset coagulopathy or death.

cluster 3 (Fig. 3c, d). Patients in cluster 3 showed higher expression of markers in signatures B, C and D than those in other clusters. Cluster 3 showed particular enrichment in expression of markers in signature B, including several innate cytokines such as IFNλ, TGFα, TSLP, IL-16, IL-23 and IL-33, and markers linked to coagulopathy, such as TPO (Fig. 3c, d).

We next ranked these parameters obtained at early time points as predictors of severe disease outcomes (Fig. 3e, Extended Data Fig. 6c). In both cases, plasma inflammatory markers were strongly associated with severe disease outcomes. For example, high levels of type I IFN (IFNα) before the first 12 DfSO correlated with longer hospital stays and death (Fig. 3e, Extended Data Fig. 6c). Moreover, patients who ultimately died of COVID-19 exhibited significantly elevated levels of IFNα, IFNλ and IL-1Ra, as well as chemokines associated with monocytes and T cell recruitment and survival such as CCL1, CLL2, macrophage colony stimulating factor (M-CSF), IL-2, IL-16 and CCL21, within the first 12 DfSO (Fig. 3e, Extended Data Fig. 6c). These analyses identify specific immunological markers that appear early in the disease and correlate strongly with poor outcomes and death.

Retrospective analysis of immune correlates

To further evaluate potential drivers of severe COVID-19 outcome in an unbiased manner, we performed unsupervised clustering analysis that included all patients and all time points using cytokines and chemokines (Fig. 4a). Notably, three main clusters of patients emerged and the distribution of patients in early time-point clusters identified in Fig. 3c matched the distribution for the all-time point analysis (Fig. 4a) in 96% of cases. Cluster 1 primarily comprised patients with moderate disease who showed improving clinical signs (Fig. 4a–d, Extended Data Fig. 7). This cluster contained only two deceased patients. Cluster 1 was characterized by low levels of inflammatory markers as well as similar or increased expression of markers in signature A′ (Fig. 4a–d), which mostly matched the signature A markers described in Fig. 3c. Clusters 2 and 3 contained patients with coagulopathy and worsened clinical progression, including most of the deceased patients (Fig. 4a–d, Extended Data Fig. 7).

Clusters 2 and 3 were driven by a set of inflammatory markers that fell into signatures B′, C′ and D′ to some extent, which overlapped highly with the 'core signature' cytokines and chemokines identified in Fig. 1 as well as with signatures B and C identified in Fig. 3c. These include type 1 immunity markers, including IL-12, chemokines linked to monocyte recruitment and IFNγ; type 2 responses, such as TSLP, chemokines linked to eosinophil recruitment, IL-4, IL-5 and IL-13; and type-3 responses, including IL-23, IL-17A and IL-22. In addition, most CRS- and inflammasome-associated cytokines were enriched in these clusters, including IL-1α, IL-1β, IL-6, IL-18 and TNF (Fig. 4a). These findings were consistent with generalized estimating equations that identified relationships between the risk of death and cytokines or immune cell populations over time (Extended Data Fig. 8). Together, these results identify groups of inflammatory and potentially protective markers that correlated with COVID-19 trajectories. The immune signatures that correlate with recovery (cluster 1) and the immune signatures that correlate with worsening diseases (cluster 2 < cluster 3) were remarkably similar whether we took a prospective (Fig. 3) or retrospective (Fig. 4) approach.

Discussion

Our longitudinal analyses of patients admitted to YNHH with COVID-19 revealed key temporal features of viral load and immune responses that distinguish disease trajectories during hospitalization. Unsupervised clustering revealed three distinct profiles that influenced the evolution and severity of COVID-19. Cluster 1, characterized by low expression of proinflammatory cytokines and enrichment in tissue repair genes, followed a disease trajectory that remained moderate and led to eventual recovery. Clusters 2 and 3 were characterized by highly elevated proinflammatory cytokines (cluster 3 being more intense), worse disease, and death. Thus, in addition to the known CRS-related pro-inflammatory cytokines, we propose these four signatures of immune response profiles that more accurately divide patients into distinct COVID-19 disease courses.

Although nasopharyngeal viral RNA levels were not significantly different between patients with moderate and severe disease at the specific time points, linear regression analyses showed a slower decline of viral loads in patients who were admitted to the ICU. Viral load was highly correlated with IFNα, IFNγ and TNF, suggesting that viral load may drive these cytokines and that interferons may not successfully control the viral replication. Moreover, many interferons, cytokines, and chemokines were elevated early in disease for patients who ultimately died of COVID-19. This finding suggests possible pathological roles associated with these host defence factors, as previous reported for patients infected with SARS-CoV-1[21].

Our comprehensive analysis of soluble plasma factors revealed broad misfiring of immune effectors in patients with COVID-19, with early predictive markers and distinct dynamics between types of immune responses among moderate and severe disease outcomes. These results suggest that late-stage pathology in COVID-19 may be driven primarily by host immune responses to SARS-CoV-2 and highlights the need for combination therapy to block other cytokines highly represented by these clusters, including inflammasome-dependent cytokines and type 2 cytokines. We observed a correlation with cytokines linked to the inflammasome pathway, which partially overlap with CRS, including IL-1β and IL-18. Indeed, it is plausible that inflammasome activation, along with a sepsis-like CRS, triggers the vascular insults and tissue pathology that are observed in patients with severe COVID-19[22].

Overall, our analyses provide a comprehensive examination of the diverse inflammatory dynamics during COVID-19 and possible contributions of distinct sets of inflammatory mediators to disease progression. This raises the possibility that early immunological interventions that target inflammatory markers that are predictive of worse disease outcome would be more beneficial than those that block late-appearing cytokines. Our disease trajectory analyses provide bases for more targeted treatment of patients with COVID-19 based on early cytokine markers, as well as therapies designed to enhance tissue repair and promote disease tolerance.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2588-y.

1. Giamarellos-Bourboulis, E. J. et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe* **27**, 992–1000.e3 (2020).
2. Zhou, Z. et al. heightened innate immune responses in the respiratory tract of COVID-19 patients. *Cell Host Microbe* **27**, 883–890.e2 (2020).
3. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
4. Mathew, D. et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* https://doi.org/10.1126/science.abc8511 (2020).
5. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
6. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
7. Chen, J. & Subbarao, K. The Immunobiology of SARS. *Annu. Rev. Immunol*. **25**, 443–472 (2007).
8. Jose, R. J. & Manuel, A. COVID-19 cytokine storm: the interplay between inflammation and coagulation. *Lancet Respir. Med*. **8**, e46–e47 (2020).
9. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest*. **130**, 2620–2629 (2020).
10. Chen, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507–513 (2020).
11. Annunziato, F., Romagnani, C. & Romagnani, S. The 3 major types of innate and adaptive cell-mediated effector immunity. *J. Allergy Clin. Immunol*. **135**, 626–635 (2015).
12. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat. Immunol*. **16**, 343–353 (2015).
13. O'Shea, J. J. & Paul, W. E. Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327**, 1098–1102 (2010).
14. Ostendorf, L. et al. Low-density granulocytes are a novel immunopathological feature in both multiple sclerosis and neuromyelitis optica spectrum disorder. *Front. Immunol*. **10**, 2725 (2019).
15. Kalfaoglu, B., Almeida-Santos, J., Adele Tye, C., Satou, Y. & Ono, M. T-cell hyperactivation and paralysis in severe COVID-19 infection revealed by single-cell analysis. Preprint at https://www.biorxiv.org/content/10.1101/2020.05.26.115923v2 (2020).
16. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e1039 (2020).
17. Kuri-Cervantes, L. et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci. Immunol*. **15**, eabd7114 (2020).
18. Gabay, C., Lamacchia, C. & Palmer, G. IL-1 pathways in inflammation and human diseases. *Nat. Rev. Rheumatol*. **6**, 232–241 (2010).
19. Wang, D. et al. The regulation of the Treg/Th17 balance by mesenchymal stem cells in human systemic lupus erythematosus. *Cell. Mol. Immunol*. **14**, 423–431 (2017).
20. Duffield, J. S., Lupher, M., Thannickal, V. J. & Wynn, T. A. Host responses in tissue repair and fibrosis. *Annu. Rev. Pathol*. **8**, 241–276 (2013).
21. Cameron, M. J. et al. Interferon-mediated immunopathological events are associated with atypical innate and adaptive immune responses in patients with severe acute respiratory syndrome. *J. Virol*. **81**, 8692–8706 (2007).
22. Yap, J. K. Y., Moriyama, M. & Iwasaki, A. Inflammasomes and pyroptosis as therapeutic targets for COVID-19. *J. Immunol*. **205**, 307–312 (2020).

Yale IMPACT Research Team

Abeer Obaid[11], Alice Lu-Culligan[1], Allison Nelson[11], Anderson Brito[3], Angela Nunez[11], Anjelica Martin[1], Annie Watkins[3], Bertie Geng[11], Chaney Kalinich[3], Christina Harden[3], Codruta Todeasa[11], Cole Jensen[3], Daniel Kim[1], David McDonald[11], Denise Shepard[11], Edward Courchaine[12], Elizabeth B. White[3], Eric Song[1], Erin Silva[11], Eriko Kudo[1], Giuseppe DeIuliis[9], Harold Rahming[11], Hong-Jai Park[11], Irene Matos[11], Jessica Nouws[11], Jordan Valdez[11], Joseph Fauver[3], Joseph Lim[13], Kadi-Ann Rose[11], Kelly Anastasio[14], Kristina Brower[3], Laura Glick[11], Lokesh Sharma[11], Lorenzo Sewanan[11], Lynda Knaggs[11], Maksym Minasyan[11], Maria Batsu[11], Mary Petrone[3], Maxine Kuang[3], Maura Nakahata[11], Melissa Campbell[8], Melissa Linehan[1], Michael H. Askenase[15], Michael Simonov[11], Mikhail Smolgovsky[11], Nicole Sonnert[1], Nida Naushad[11], Pavithra Vijayakumar[11], Rick Martinello[4], Rupak Datta[4], Ryan Handoko[11], Santos Bermejo[11], Sarah Prophet[16], Sean Bickerton[12], Sofia Velazquez[15], Tara Alpert[4], Tyler Rice[1], William Khoury-Hanold[1], Xiaohua Peng[11], Yexin Yang[1], Yiyun Cao[1] & Yvette Strong[11]

[11]Yale University School of Medicine, New Haven, CT, USA. [12]Department of Biochemistry and of Molecular Biology, Yale University School of Medicine, New Haven, CT, USA. [13]Yale Viral Hepatitis Program, Yale University School of Medicine, New Haven, CT, USA. [14]Yale Center for Clinical Investigation, Yale University School of Medicine, New Haven, CT, USA. [15]Department of Neurology, Yale University School of Medicine, New Haven, CT, USA. [16]Department of Molecular, Cellular and Developmental Biology, Yale University School of Medicine, New Haven, CT, USA.

Methods

Ethics statement

This study was approved by Yale Human Research Protection Program Institutional Review Boards (FWA00002571, protocol ID 2000027690). Informed consent was obtained from all enrolled patients and healthcare workers.

Patients

One-hundred and thirty-five patients admitted to YNHH with COVID-19 between 18 March 2020and 5 May 2020 were included in this study. No statistical methods were used to predetermine sample size. Nasopharyngeal swabs were collected as described[23], approximately every four days, for SARS-CoV-2 RT–qPCR analysis where clinically feasible. Paired whole blood for flow cytometry analysis was collected simultaneously in sodium heparin-coated vacutainers and kept on gentle agitation until processing. All blood was processed on the day of collection. Patients were scored for COVID-19 disease severity through review of electronic medical records (EMR) at each longitudinal time point. Scores were assigned by a clinical infectious disease physician according to a custom-developed disease severity scale. Moderate disease status (clinical score 1–3) was defined as: SARS-CoV-2 infection requiring hospitalization without supplementary oxygen (1); infection requiring non-invasive supplementary oxygen (<3 l/min to maintain SpO$_2$ >92%) (2); and infection requiring non-invasive supplementary oxygen (>3 l/min to maintain SpO$_2$ >92%, or >2 l/min to maintain SpO$_2$ >92% and had a high-sensitivity C-reactive protein (CRP) >70) and received tocilizumab). Severe disease status (clinical score 4 or 5) was defined as infection meeting all criteria for clinical score 3 and also requiring admission to the ICU and >6 l/min supplementary oxygen to maintain SpO$_2$ >92% (4); or infection requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO) in addition to glucocorticoid or vasopressor administration (5). Clinical score 6 was assigned for deceased patients. Of note, the use of tocilizumab can increase circulating levels of IL-6 by inhibiting IL-6Rα-mediated degradation. Analysis of our cohort indicate higher plasma levels of IL-6 in patients with either moderate or severe disease who received tocilizumab treatment (Extended Data Fig. 1d).

For all patients, days from symptom onset were estimated as follows: (1) highest priority was given to explicit onset dates provided by patients; (2) next highest priority was given to the earliest reported symptom by a patient; and (3) in the absence of direct information regarding symptom onset, we estimated a date through manual assessment of the electronic medical record (EMRs) by an independent clinician. Demographic information was aggregated through a systematic and retrospective review of patient EMRs and was used to construct Extended Data Table 1. Symptom onset and aetiology were recorded through standardized interviews with patients or patient surrogates upon enrollment in our study, or alternatively through manual EMR review if no interview was possible owing to clinical status. The clinical data were collected using EPIC EHR and REDCap 9.3.6 software. At the time of sample acquisition and processing, investigators were unaware of the patients' conditions. Blood acquisition was performed and recorded by a separate team. Information about patients' conditions was not available until after processing and analysis of raw data by flow cytometry and ELISA. A clinical team, separate from the experimental team, performed chart reviews to determine relevant statistics. Cytokines and FACS analyses were performed blinded. Patients' clinical information and clinical score coding were revealed only after data collection.

Viral RNA measurements

RNA concentrations were measured from nasopharyngeal samples by RT–qPCR as previously described[23]. In brief, total nucleic acid was extracted from 300 µl of viral transport medium (nasopharyngeal swab) using the MagMAX Viral/Pathogen Nucleic Acid Isolation kit (ThermoFisher Scientific) with a modified protocol and eluted into 75 µl elution buffer.

To detect SARS-CoV-2 RNA, we tested 5 µl RNA 371 template as previously described[24], using the US CDC real-time RT–qPCR primer/probe sets for 2019-nCoV_N1, 2019-nCoV_N2, and the human RNase P (RP) as an extraction control. Virus RNA copies were quantified using a tenfold dilution standard curve of RNA transcripts that we previously generated[24]. The lower limit of detection for SARS-CoV-2 genomes assayed by qPCR in nasopharyngeal specimens was established as described[24]. In addition to a technical detection threshold, we also used a clinical referral threshold (detection limit) to either: (1) refer asymptomatic HCWs for diagnostic testing at a CLIA-approved laboratory; or (2) cross-validate results from a CLIA-approved laboratory for SARS-CoV-2 qPCR-positive individuals upon study enrollment. Individuals above the technical detection threshold, but below the clinical referral threshold, were considered SARS-CoV-2 positive for the purposes of our research.

Isolation of patient plasma

Plasma samples were collected after centrifugation of whole blood at 400*g* for 10 min at room temperature (RT) without brake. The undiluted serum was then transferred to 15-ml polypropylene conical tubes, and aliquoted and stored at −80 °C for subsequent analysis.

Cytokine and chemokine measurements

Patient serum was isolated as before and aliquots were stored at −80 °C. Sera were shipped to Eve Technologies (Calgary, Alberta, Canada) on dry ice, and levels of cytokines and chemokines were measured using the Human Cytokine Array/Chemokine Array 71-403 Plex Panel (HD71). All samples were measured upon the first thaw.

Isolation of PBMCs

PBMCs were isolated from heparinized whole blood using Histopaque (Sigma-Aldrich, #10771-500ML) density gradient centrifugation in a biosafety level 2+ facility. After isolation of undiluted serum, blood was diluted 1:1 in room temperature PBS, layered over Histopaque in a SepMate tube (StemCell Technologies; #85460) and centrifuged for 10 min at 1,200*g*. The PBMC layer was isolated according to the manufacturer's instructions. Cells were washed twice with PBS before counting. Pelleted cells were briefly treated with ACK lysis buffer for 2 min and then counted. Percentage viability was estimated using standard Trypan blue staining and an automated cell counter (Thermo-Fisher, #AMQAX1000).

Flow cytometry

Antibody clones and vendors were as follows: BB515 anti-hHLA-DR (G46-6) (1:400) (BD Biosciences), BV785 anti-hCD16 (3G8) (1:100) (BioLegend), PE-Cy7 anti-hCD14 (HCD14) (1:300) (BioLegend), BV605 anti-hCD3 (UCHT1) (1:300) (BioLegend), BV711 anti-hCD19 (SJ25C1) (1:300) (BD Biosciences), AlexaFluor647 anti-hCD1c (L161) (1:150) (BioLegend), biotin anti-hCD141 (M80) (1:150) (BioLegend), PE-Dazzle594 anti-hCD56 (HCD56) (1:300) (BioLegend), PE anti-hCD304 (12C2) (1:300) (BioLegend), APCFire750 anti-hCD11b (ICRF44) (1:100) (BioLegend), PerCP/Cy5.5 anti-hCD66b (G10F5) (1:200) (BD Biosciences), BV785 anti-hCD4 (SK3) (1:200) (BioLegend), APCFire750 or PE-Cy7 or BV711 anti-hCD8 (SK1) (1:200) (BioLegend), BV421 anti-hCCR7 (G043H7) (1:50) (BioLegend), AlexaFluor 700 anti-hCD45RA (HI100) (1:200) (BD Biosciences), PE anti-hPD1 (EH12.2H7) (1:200) (BioLegend), APC anti-hTIM3 (F38-2E2) (1:50) (BioLegend), BV711 anti-hCD38 (HIT2) (1:200) (BioLegend), BB700 anti-hCXCR5 (RF8B2) (1:50) (BD Biosciences), PE-Cy7 anti-hCD127 (HIL-7R-M21) (1:50) (BioLegend), PE-CF594 anti-hCD25 (BC96) (1:200) (BD Biosciences), BV711 anti-hCD127 (HIL-7R-M21) (1:50) (BD Biosciences), BV421 anti-hIL17a (N49-653) (1:100) (BD Biosciences), AlexaFluor 700 anti-hTNFa (MAb11) (1:100) (BioLegend), PE or APC/Fire750 anti-hIFNγ (4S.B3) (1:60) (BioLegend), FITC anti-hGranzymeB (GB11)

(1:200) (BioLegend), AlexaFluor 647 anti-hIL-4 (8D4-8) (1:100) (BioLegend), BB700 anti-hCD183/CXCR3 (1C6/CXCR3) (1:100) (BD Biosciences), PE-Cy7 anti-hIL-6 (MQ2-13A5) (1:50) (BioLegend), PE anti-hIL-2 (5344.111) (1:50) (BD Biosciences), BV785 anti-hCD19 (SJ25C1) (1:300) (BioLegend), BV421 anti-hCD138 (MI15) (1:300) (BioLegend), AlexaFluor700 anti-hCD20 (2H7) (1:200) (BioLegend), AlexaFluor 647 anti-hCD27 (M-T271) (1:350) (BioLegend), PE/Dazzle594 anti-hIgD (IA6-2) (1:400) (BioLegend), PE-Cy7 anti-hCD86 (IT2.2) (1:100) (BioLegend), APC/Fire750 anti-hIgM (MHM-88) (1:250) (BioLegend), BV605 anti-hCD24 (ML5) (1:200) (BioLegend), BV421 anti-hCD10 (HI10a) (1:200) (BioLegend), BV421 anti-CDh15 (SSEA-1) (1:200) (BioLegend), AlexaFluor 700 Streptavidin (1:300) (ThermoFisher), BV605 Streptavidin (1:300) (BioLegend). In brief, freshly isolated PBMCs were plated at 1–2 × 10^6 cells per well in a 96-well U-bottom plate. Cells were resuspended in Live/Dead Fixable Aqua (ThermoFisher) for 20 min at 4 °C. Following a wash, cells were blocked with Human TruStan FcX (BioLegend) for 10 min at RT. Cocktails of desired staining antibodies were added directly to this mixture for 30 min at RT. For secondary stains, cells were first washed and supernatant aspirated; then to each cell pellet a cocktail of secondary markers was added for 30 min at 4 °C. Prior to analysis, cells were washed and resuspended in 100 μl 4% PFA for 30 min at 4 °C. For intracellular cytokine staining following stimulation, cells were resuspended in 200 μl cRPMI (RPMI-1640 supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 mg/ml streptomycin, 1 mM sodium pyruvate, and 50 μM 2-mercaptoethanol) and stored at 4 °C overnight. Subsequently, these cells were washed and stimulated with 1× Cell Stimulation Cocktail (eBioscience) in 200 μl cRPMI for 1 h at 37 °C. Fifty microlitres of 5× Stimulation Cocktail (plus protein transport 442 inhibitor) (eBioscience) was added for an additional 4 h of incubation at 37 °C. Following stimulation, cells were washed and resuspended in 100 μl 4% PFA for 30 min at 4 °C. To quantify intracellular cytokines, these samples were permeabilized with 1× permeabilization buffer from the FOXP3/Transcription Factor Staining Buffer Set (eBioscience) for 10 min at 4 °C. All subsequent staining cocktails were made in this buffer. Permeabilized cells were then washed and resuspended in a cocktail containing Human TruStan FcX (BioLegend) for 10 min at 4 °C. Finally, intracellular staining cocktails were added directly to each sample for 1 h at 4 °C. Following this incubation, cells were washed and prepared for analysis on an Attune NXT (ThermoFisher). Data were analysed using FlowJo software version 10.6 software (Tree Star). The specific sets of markers used to identify each subset of cells are summarized in Extended Data Fig. 9.

Statistical analysis

Patients and their analysed features were clustered using the K-means algorithm. Heat maps were created using the ComplexHeatmap package[25]. The optimum number of clusters was determined by using the silhouette coefficient analysis, available with the NBClust and factoextra packages[26]. Before data visualization, each feature was scaled and centred. Multiple group comparisons were analysed by running both parametric (ANOVA) and non-parametric (Kruskal–Wallis) statistical tests with Dunn's and Tukey's post hoc tests. Mutual information analyses were performed using the Caret R package and visualized using ggplot2. Multiple correlation analysis was performed by computing Spearman's coefficients with the Hmisc package for R and visualized with corrplot by only showing correlations with $P < 0.05$. For generalized linear models (GLM), we calculated the incident risk ratio (IRR) by conducting a Poisson regression with a log link and robust variance estimation; this value approximates the risk ratio estimated by a log-linear model. For generalized estimating equation (GEE) models, we calculated the incidence risk ratio (IRR) in the same way as for non-GEE GLM models, assuming an independent correlation structure. All models controlled for participant sex and age.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All the background information on HCWs, clinical information for patients, and raw data used in this study are included in Supplementary Table 1. Additionally, all of the raw fcs files for the flow cytometry analysis are available at ImmPort (https://www.immport.org/shared/home; study ID SDY1655).

23. Wyllie, A. L. et al. Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients than nasopharyngeal swabs. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.16.20067835v1 (2020).
24. Vogels, C. B. F. et al. Analytical sensitivity and efficiency comparisons of SARS-COV-2 qRT-PCR primer-probe sets. *Nat. Microbiol*. https://doi.org/10.1038/s41564-020-0761-6 (2020).
25. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
26. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw*. **61**, 1–36 (2014).

Acknowledgements We thank M. Linehan for technical and logistical assistance, and A. Wang, A. Ring, C. Wilen and D. Mucida for discussions. This work was supported by the Women's Health Research at Yale Pilot Project Program (A.I.), Fast Grant from Emergent Ventures at the Mercatus Center, Mathers Foundation, and the Ludwig Family Foundation, the Department of Internal Medicine at the Yale School of Medicine, Yale School of Public Health and the Beatrice Kleinberg Neuwirth Fund. IMPACT received support from the Yale COVID-19 Research Resource Fund. A.I. is an Investigator of the Howard Hughes Medical Institute. C.L. is a Pew Latin American Fellow. P.W. is supported by Gruber Foundation and the NSF. B.I. is supported by NIAID 2T32AI007517-16. C.B.F.V. is supported by NOW Rubicon 019.181EN.004.

Author contributions A.I.K. and A.I. conceived the study. C.L., P.W., J.K., J.S., J.E.O. S.M., H.W. and T.M. defined parameters, collected and processed patient PBMC samples and analysed data. T.B.R.C. performed bioinformatic analysis. B.I., J.K. T.T. and C.D.O. collected epidemiological and clinical data. A.L.W., C.B.F.V., I.M.O., R.E., S.L., P.L., A.V., A.P. and M.T. performed the virus RNA concentration assays. N.D.G. supervised the virus RNA concentration assays. A.C.-M., M.C.M and A.J.M. processed and stored patient specimens. J.B.F., C.D.C. M.C. and S.F. assisted in patient and HCW recruitment. W.L.S. supervised clinical data management. M.S., M.K.E. and S.B.O. carried out statistical analyses. A.C.S. and R.M. contributed personnel, equipment and insights. C.L. and A.I. drafted the manuscript. All authors helped to edit the manuscript. A.I. and R.H. secured funds. A.I. and S.B.O. supervised the project.

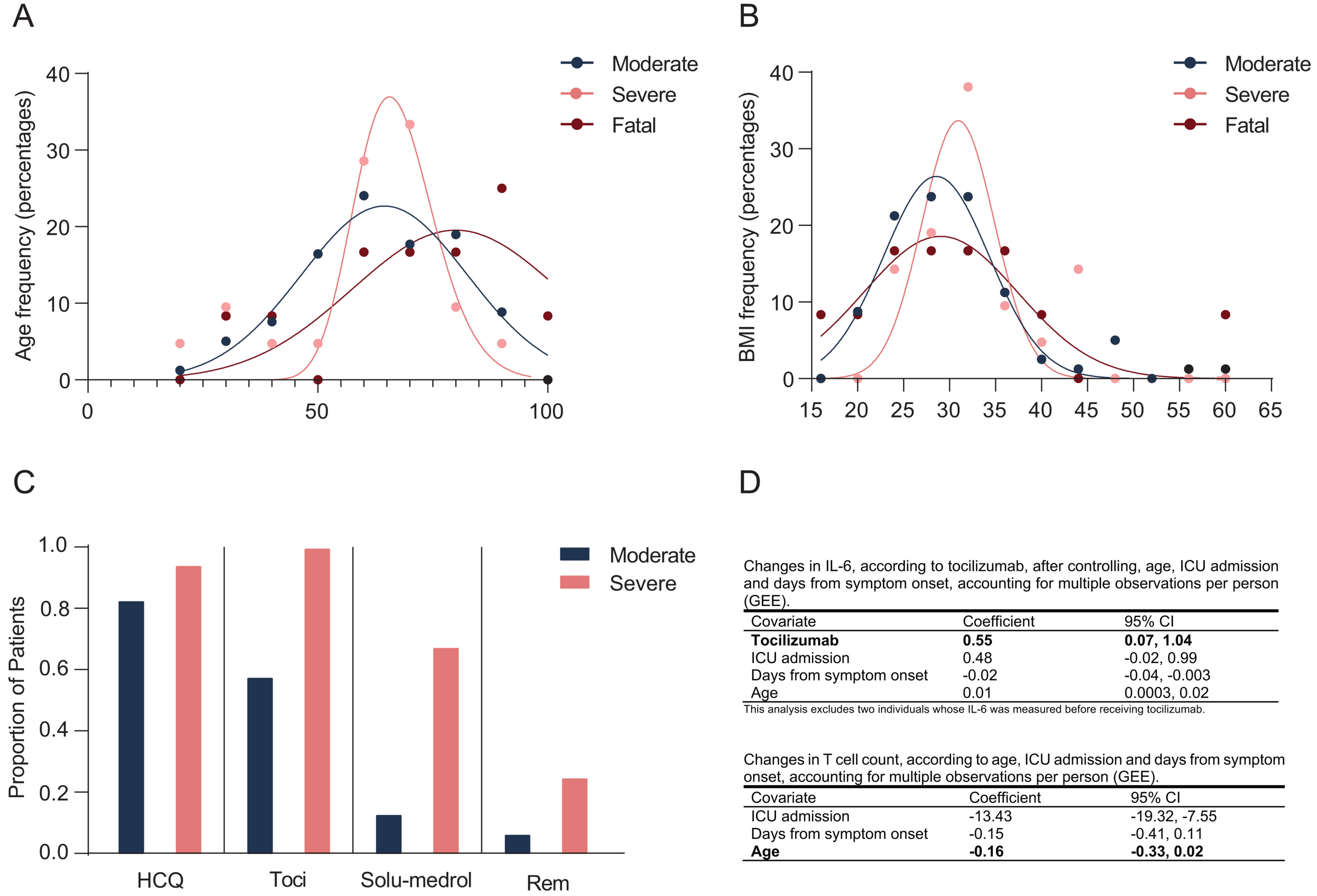
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2588-y.

Correspondence and requests for materials should be addressed to A.I.

Peer review information *Nature* thanks Petter Brodin, Malik Peiris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Changes in IL-6, according to tocilizumab, after controlling, age, ICU admission and days from symptom onset, accounting for multiple observations per person (GEE).

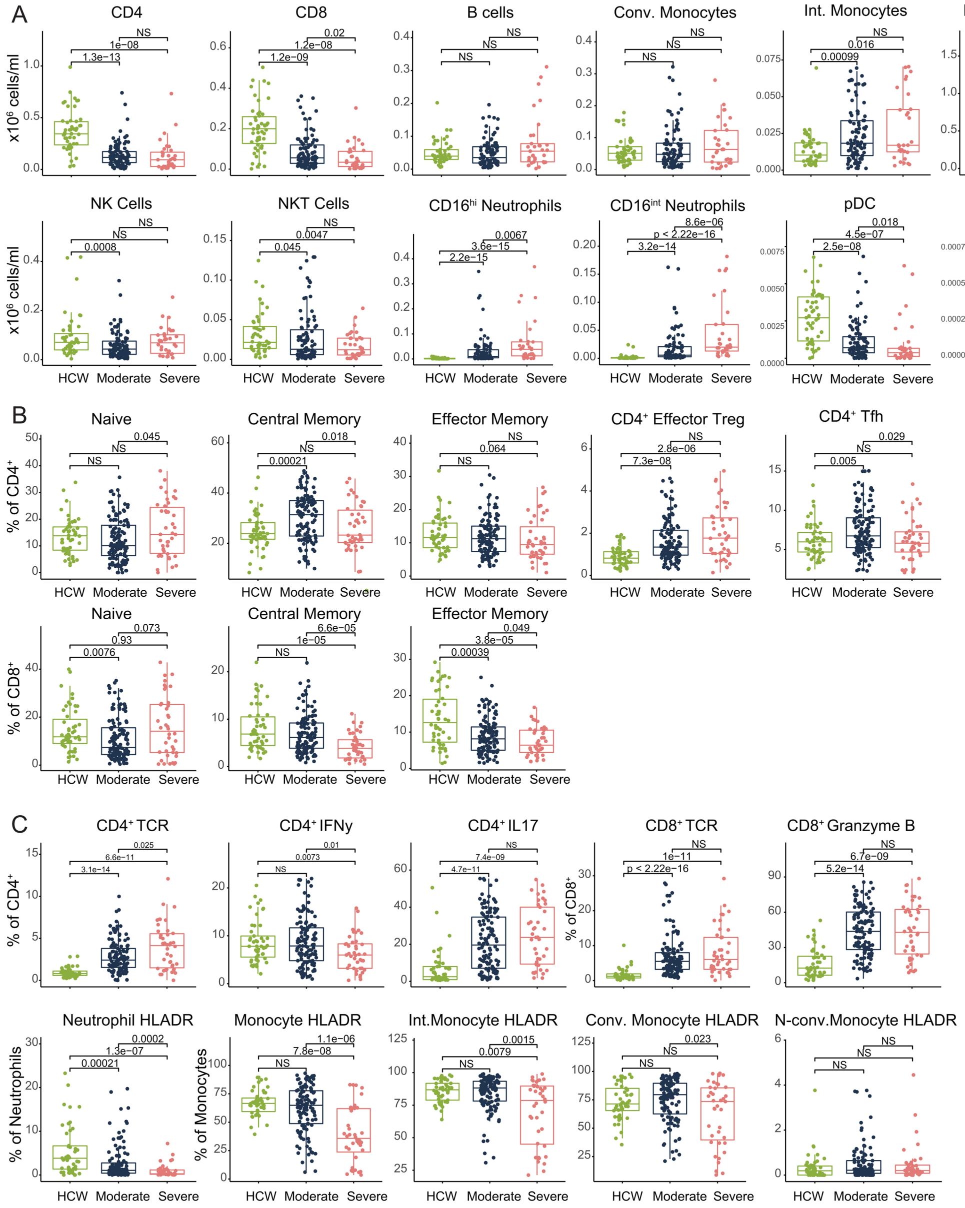
Covariate	Coefficient	95% CI
Tocilizumab	**0.55**	**0.07, 1.04**
ICU admission	0.48	-0.02, 0.99
Days from symptom onset	-0.02	-0.04, -0.003
Age	0.01	0.0003, 0.02

This analysis excludes two individuals whose IL-6 was measured before receiving tocilizumab.

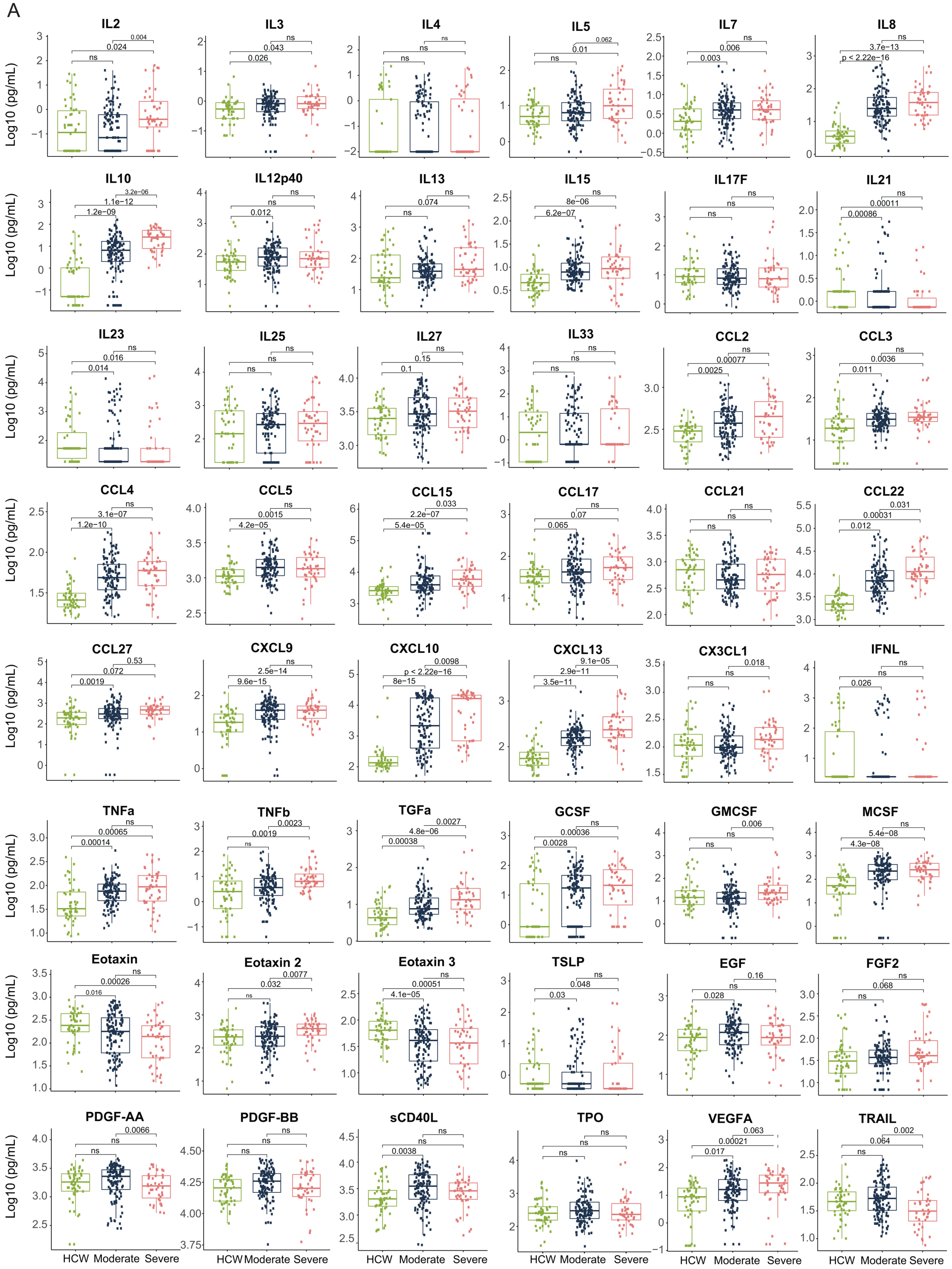
Changes in T cell count, according to age, ICU admission and days from symptom onset, accounting for multiple observations per person (GEE).

Covariate	Coefficient	95% CI
ICU admission	-13.43	-19.32, -7.55
Days from symptom onset	-0.15	-0.41, 0.11
Age	**-0.16**	**-0.33, 0.02**

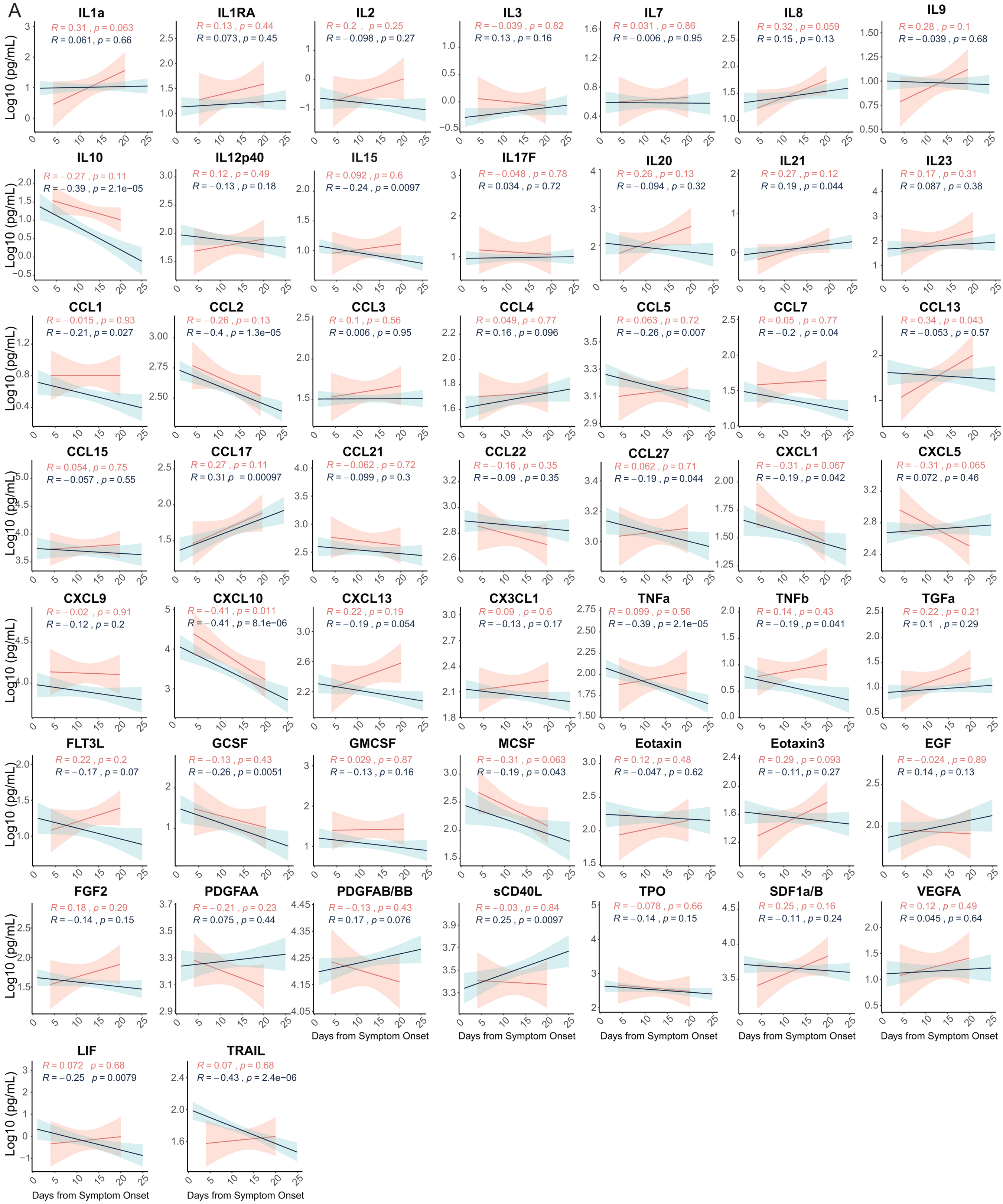
Extended Data Fig. 1 | Age and BMI cohort distributions and Select Medications distributions. a, **b**, Aggregated ages (**a**) and BMIs (**b**) were collected for patients with moderate, severe, and fatal COVID-19 and relative frequency histograms generated for comparison across disease sub-groups. Gaussian and lognormal distributions were fit through least squares regression and compared for goodness of fit through differential Akaike information criterion (AICc) comparison. All distributions were best described by a Gaussian model except for age in the 'severe' disease category, which was best modelled by a lognormal distribution. **c**, Proportion of patients admitted to YNHH receiving hydroxycholorquine (HCQ), tocilizumab (Toci), methylprednisolone (Solu-medrol), and remdesivir (Rem) are shown, stratified by disease severity. **d**, Medication and age adjustments for IL-6 and T cell count.



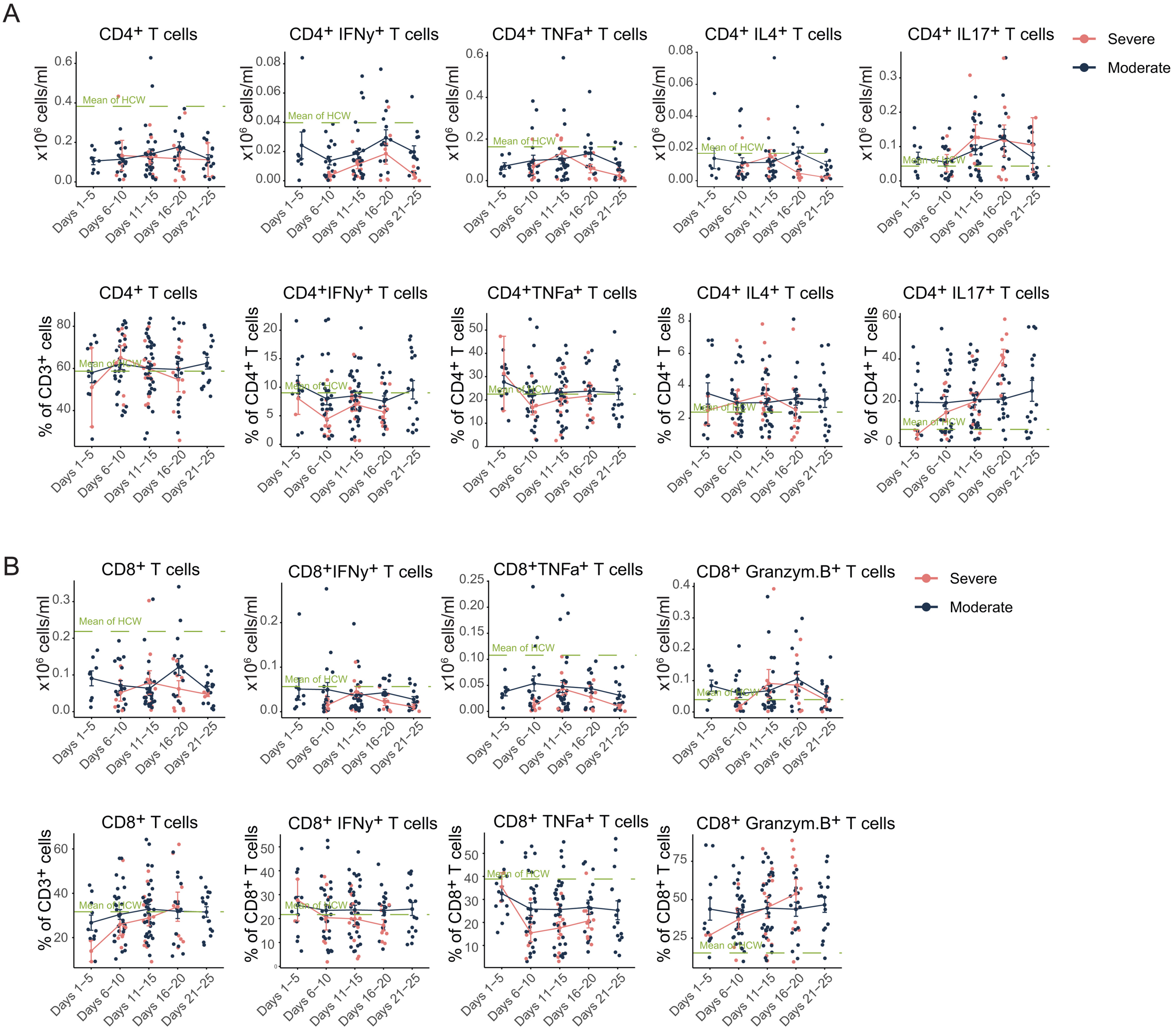
Extended Data Fig. 2 | Overview of cellular immune changes in COVID-19 patients. a, **b**, Immune cell subsets of interest, plotted as a concentration of millions of cells per millilitre of blood (**a**) or as a percentage of a parent population (**b**). **c**, Phenotyping to TCR-activated T cells, cytokine-secreting T cells, and HLA-DR expression within monocytes and neutrophils. Each dot represents a separate time point per subject (HCW, *n* = 49; Moderate, *n* = 114; Severe, *n* = 41). For all boxplots, the centre is drawn through the median of the measurement, and the lower and upper bounds of the box correspond to the first and third percentile. Whiskers beyond these points denote 1.5 × the interquartile range. *P* values were determined by two-sided, Wilcoxon rank-sum test.



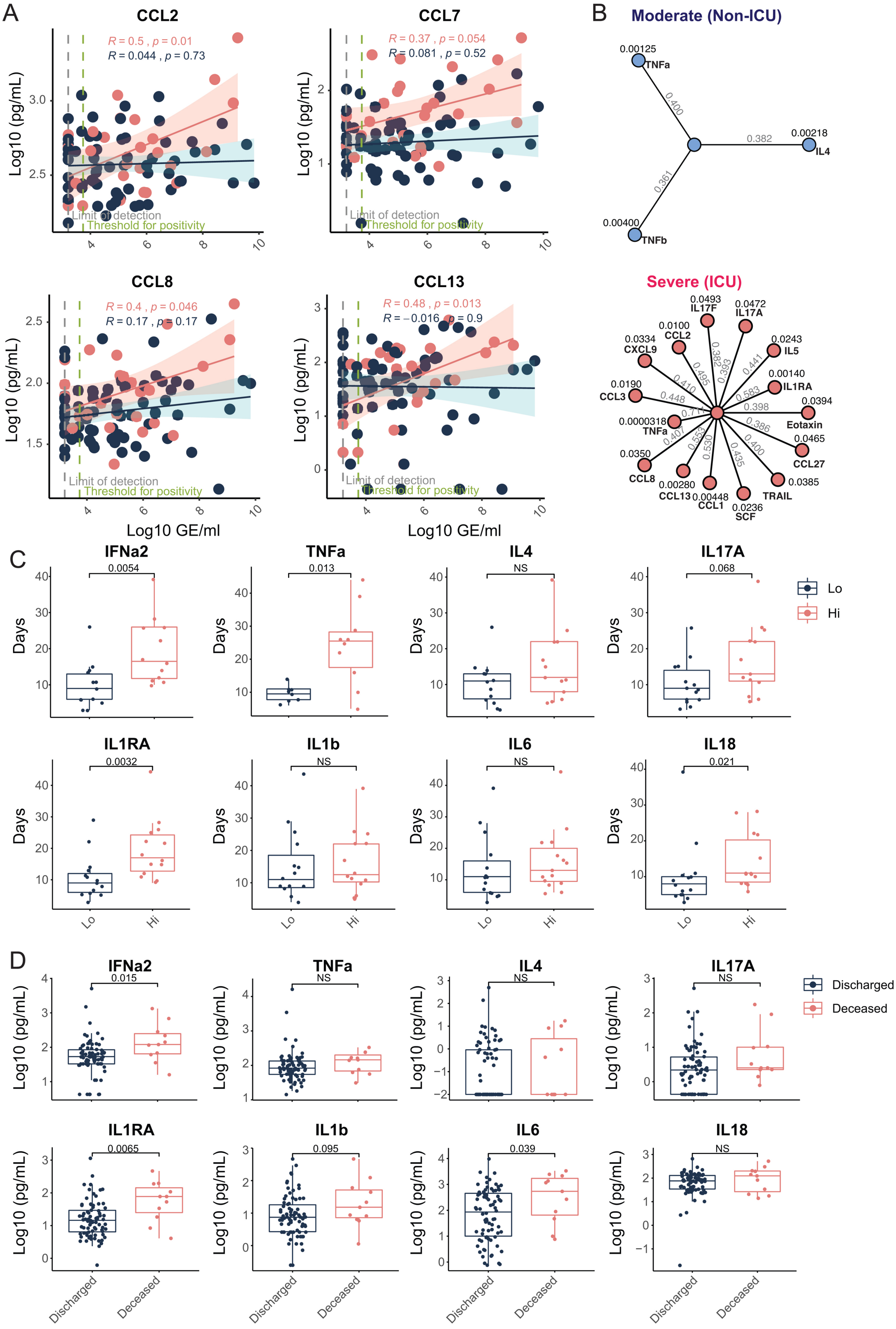
Extended Data Fig. 3 | Overview cytokine and chemokines profiles of COVID-19 patients. a, Quantification of cytokines in the periphery plotted as $\log_{10}$-transformed concentrations. Each dot represents a separate time point per subject (HCW, *n* = 47; Moderate, *n* = 124; Severe, *n* = 45). For all boxplots, the centre is drawn through the median of the measurement, while the lower and upper bounds of the box correspond to the first and third percentile. Whiskers beyond these points denote 1.5 × the interquartile range. *P* values were determined by two-sided, Wilcoxon rank-sum test.



Extended Data Fig. 4 | Longitudinal cytokines and chemokines of COVID-19 patients. a, Quantification of cytokines plotted as $\log_{10}$-transformed concentration over time according to the days of symptom onset for patients with moderate disease ($n = 112$) or severe disease ($n = 39$). The dotted green line represents the mean measurement from uninfected HCWs. Regression lines are indicated by the dark blue (moderate) or red (severe) solid lines. Associated, Pearson's correlation coefficients and linear regression significance are in pink (moderate) or dark blue (severe). 95% confidence intervals for the regression lines are denoted by the pink (moderate) or dark blue (severe) filled areas.

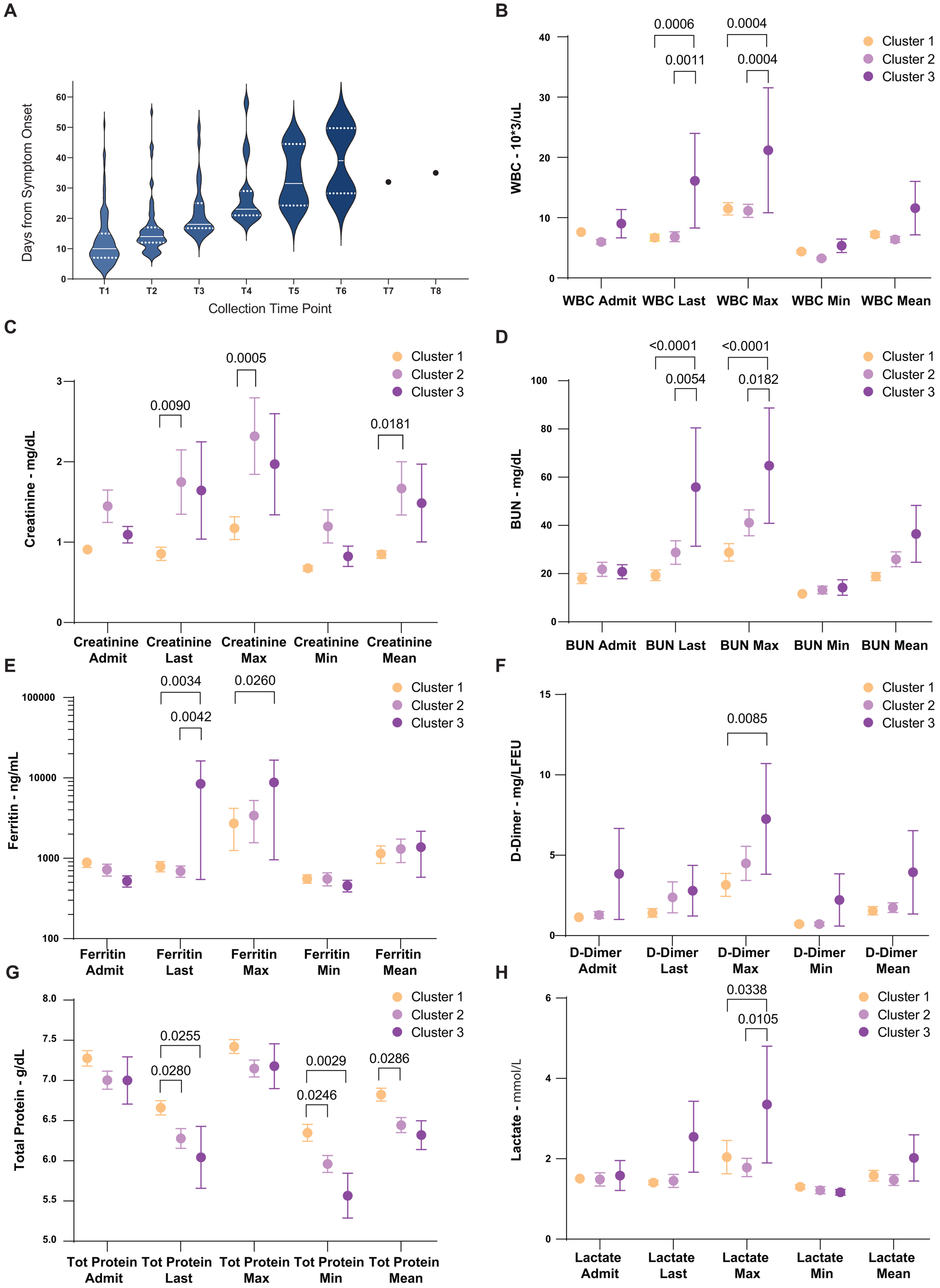


Extended Data Fig. 5 | T cell immune profiles in moderate and severe patients. a, **b**, CD4$^+$ (**a**) and CD8$^+$ (**b**) T cell populations of interest, plotted as a percentage of parent populations, over time according to the days following symptom onset for patients with moderate disease ($n = 118$) or severe disease ($n = 41$). Each dot represents a distinct patient and time point arranged by intervals of five days until 25 days. Dark blue or pink lines pass through the mean of each measurement at the specified time interval; error bars at this intersection denote s.e.m. The dotted green line represents the mean measurement from uninfected HCWs.



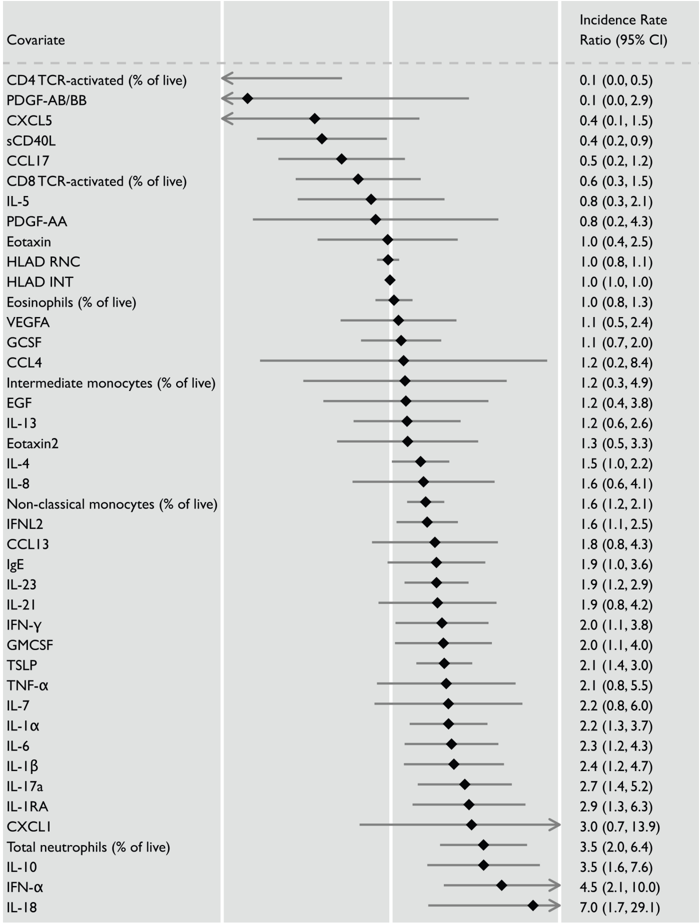
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Early cytokine profile distinguishes moderate and severe outcomes. a, Quantification of $\log_{10}$-transformed cytokine concentrations plotted continuously with NP viral load (expressed as $\log_{10}$ genomic equivalents (GE)/ml) per within an individual patient and time point. Regression lines are indicated by the dark blue (moderate) or red (severe) solid lines for patients with moderate disease (*n* = 112) or severe disease (*n* = 39), respectively. Associated Pearson's correlation coefficients, and linear regression significance are in pink (moderate) or dark blue (severe). 95% confidence intervals for the regression lines are denoted by the pink (moderate) or dark blue (severe) filled areas. **b**, Correlation map of highly correlated cytokines with NP viral load in patients with moderate (blue) or severe disease (red). Pearson's correlation coefficients are indicated in grey, connecting the central node, NP viral load, with peripheral nodes; *P* values for each correlation are indicated above each peripheral node. **c**, Length of hospital stay plotted per patient against an individual's baseline plasma cytokine measurements (<12 days from symptom onset), which were grouped according to high or low expression (>0.5 $\log_{10}$-transformed difference): IFNa2 (Hi:12, Lo:13), TNFa (Hi:6, Lo:4), IL4 (Hi:7, Lo:11), IL4 (Hi:8, Lo:6), IL1RA (Hi:8, Lo:7), IL1b (Hi:11, Lo:5), IL6 (Hi:8, Lo:7), IL18 (Hi:5, Lo:5). **d**, Baseline plasma cytokine measurements for each patient who was either discharged from the hospital (*n* = 83) or expired during treatment for COVID-19 (*n* = 11). For all boxplots, the centre is drawn through the median of the measurement, while the lower and upper bounds of the box correspond to the first and third percentile. Whiskers beyond these points denote 1.5 × the interquartile range. *P* values were determined by two-sided, Wilcoxon rank-sum test.

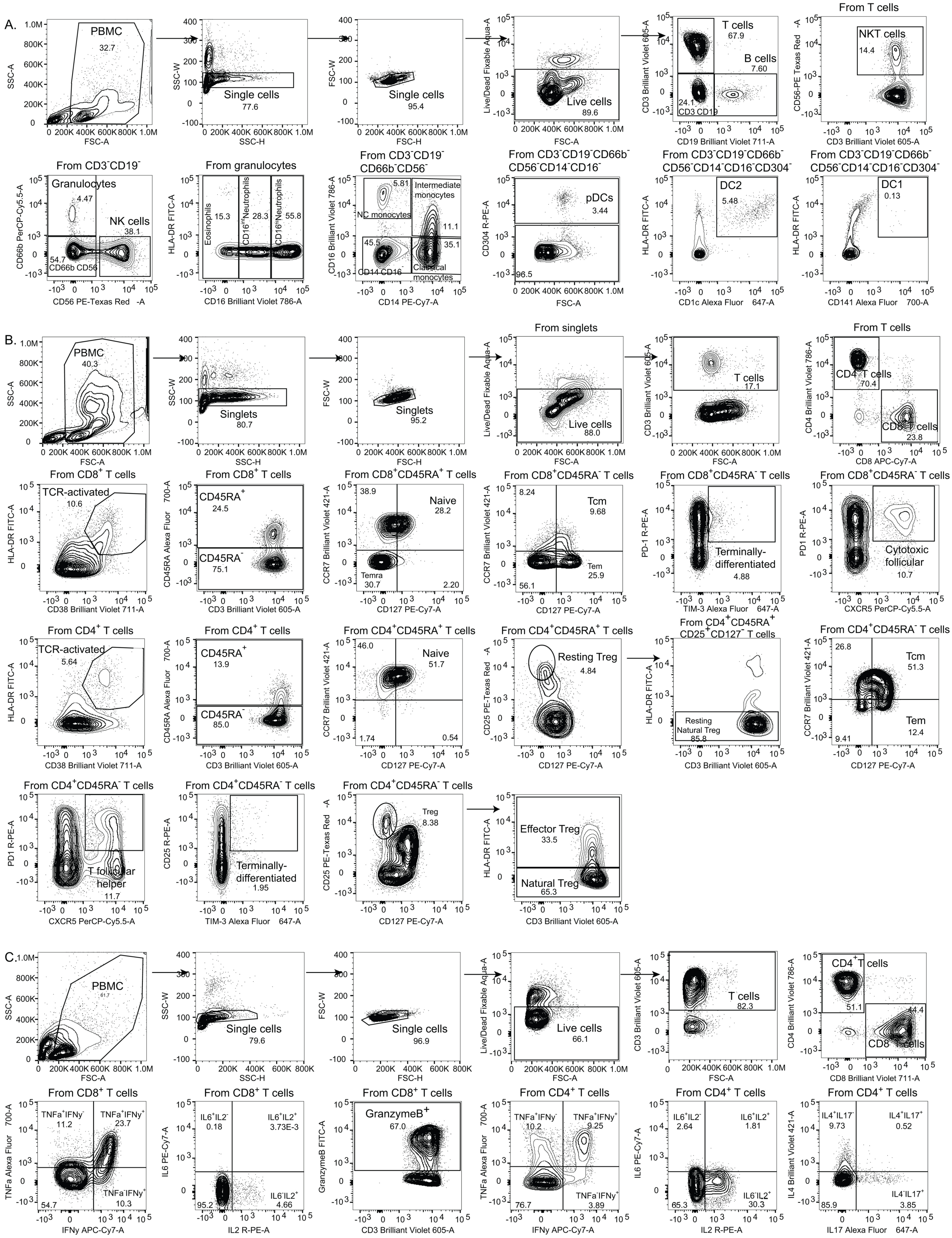


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Distribution of days from symptom onset stratified by collection time point and select cluster clinical data. **a**, Correlation of days from symptom onset and samples collection time points. Violin plots comparing the distributions of days from symptom for each patient ordered by sequential IMPACT study time points (1–8). Study time points 7 and 8 are represented by discrete points for the single patient collected at each. Violin plots display median values (solid line) and associated quartiles (dashed lines). T1–8 (time point 1 to 8). **b**–**h**, Aggregated clinical data for patients in clusters 1–3. Displayed are laboratory values at time of admission to YNHH ("admit"); last recorded values from duration of admission ("last"); maximum recorded values from duration of admission ("max"); minimum recorded values from duration of admission ("min"); and average recorded values for duration of admission ("mean"). Scatter plots show cluster means with s.e.m. plotted above and below. Clusters were subsequently compared using ordinary two-way ANVOA and post hoc pairwise comparisons are identified where significant (adjusted *P* values displayed, Tukey's method for multiple comparisons).



Extended Data Fig. 8 | Risk of death according to biomarkers levels. Forest plots comparing the risk of death among ill patients. Each effect estimate represents an individual regression estimate with a Poisson family, log link, and robust variance estimation; each model accounts for repeated measures within one individual through the use of generalized estimating equations (GEE). Measurements are divided into three time-periods: 0–11 days after symptom onset, 12–19 days after symptom onset, and ≥20 days after symptom onset. If an individual had more than one measurement of a biomarker during any particular time period, we used the average of all values. Each model controls for participant age and gender.



Extended Data Fig. 9 | Gating strategies. Gating strategies are shown for the key cell populations described in Figs. 1b, c, 2d–f, and in Extended Data Figures. **a**, Leukocyte gating strategy to identify lymphocytes, granulocytes, monocytes, pDCs, and cDCs in Figs. 1b, c, 2d–f and Extended Data Fig. 2a. **b**, T cell surface staining gating strategy to identify CD4 and CD8 T cells, TCR-activated T cells, terminally-differentiated T cells, and additional subsets as shown in Extended Data Fig. 2b. **c**, Intracellular T cell gating strategy to identify CD4 and/or CD8 T cells secreting TNF, IFNγ, IL-6, IL-2, granzyme B, IL-4, and/or IL-17 in Extended Data Figs. 2c, 5a, b.

Extended Data Table 1 | Basic demographics for IMPACT cohort

	Moderate COVID-19	Severe COVID-19	Relative Risk (95% CI); [*; p-value]	Total
Number	70.8% (80 / 113)	29.2% (33 / 113)		113
Age (years)	62.66 ± 16.1	63.67 ± 19.3	[n.s.]	62.96 ± 17.0
Sex				
Male	45% (36 / 80)	48.48% (16 / 33)	1.07 (.70 - 1.65)	46.02% (52 / 113)
Female	55% (44 / 80)	51.52% (17 / 33)	.94 (.64 - 1.38)	53.98% (61 / 113)
Ethnicity				
American Indian / Alaskan Native	0% (0 / 80)	0% (0 / 33)	--	0% (0 / 113)
Asian	1.25% (1 / 80)	0% (0 / 33)	--	0.88% (1 / 113)
Black / African American	27.5% (22 / 80)	33.33% (11 /33)	1.21 (.67 - 2.21)	29.2% (33 / 113)
Native Hawaiian / Pacific Islander	0% (0 / 80)	0% (0 / 33)	-	0% (0 / 113)
White	53.75% (43 / 80)	54.55% (18 / 33)	1.01 (.70 - 1.47)	53.98% (61 / 113)
Hispanic	12.5% (10 / 80)	12.12% (4 / 33)	.97 (.33 - 2.87)	12.39% (14 / 113)
Multiple	0% (0 / 80)	0% (0 / 33)	--	0% (0 / 113)
Unknown	5% (4 / 80)	0% (0 / 33)	n.c.	3.54% (4 / 113)
BMI				
<18.5	0% (0 / 80)	6.06% (2 / 33)	n.c.	1.77% (2 / 113)
18.5-24.9	21.25% (17 / 80)	9.09% (3 / 33)	.43 (.13 - 1.36)	17.7% (20 / 113)
25.0-29.9	32.5% (26 / 80)	24.24% (8 / 33)	.75 (.38 - 1.47)	30.09% (34 / 113)
30-35	27.5% (22 / 80)	30.3% (10 / 33)	1.10 (.59 - 2.06)	28.32% (32 / 113)
>35	18.75% (15 / 80)	30.3% (10 / 33)	1.62 (.81 - 3.22)	22.12% (25 / 113)
COVID Risk Factors				
None	27.5% (22 / 80)	30.3% (10 / 33)	1.10 (.59 - 2.06)	28.32% (32 / 113)
Cancer Treatment within 1 year	7.5% (6 / 80)	15.15% (5 / 33)	2.02 (.66 - 6.16)	9.73% (11 / 113)
Chronic Heart Disease	27.5% (22 / 80)	24.24% (8 / 33)	.88 (.44 - 1.78)	26.55% (30 / 113)
Hypertension	53.75% (43 / 80)	48.48% (16 / 33)	.90 (.60 - 1.35)	52.21% (59 / 113)
Chronic Lung Disease (asthma, COPD, ILD)	26.25% (21 / 80)	18.18% (6 / 33)	.69 (.31 - 1.56)	23.89% (27 / 113)
Immunosupression	11.25% (10 / 80)	6.06% (2 / 33)	.52 (.12 - 2.29)	9.73% (12 / 113)
Solid Organ Transplant	6.25% (4 / 80)	3.03% (1 / 33)	.60 (.07 - 5.16)	4.42% (5 / 113)
HIV^+ (with anti-viral treatment; CD4 > 400)	2.5% (2/80)	0% (0 / 33)	n.c.	1.77% (2 / 113)
Other (Multiple Sclerosis, Rheumatoid Arthritis, Scleroderma, Cirrhosis)	3.75% (3 / 80)	3.03% (1 / 33)	1.21 (.11 - 12.91)	3.54% (4 / 113)
Presenting Symptoms				
Headache	56.9% (33 / 58)	47.37% (9 / 19)	.83 (.49 - 1.41)	54.55% (42 / 77)
Objective Fever (> 100.3 °F / 37.9 °C)	64.29% (36 / 56)	65% (13 / 20)	1.01 (.69 - 1.47)	64.47% (49 / 76)
Cough	77.19% (44 / 57)	65% (13 / 20)	.84 (.59 - 1.20)	74.03% (57 / 77)
Dyspnea	64.41% (38 / 59)	75% (15 / 20)	1.16 (.85 - 1.60)	67.09% (53 / 79)
Rhinorrhea	30.36% (17 / 56)	35.29% (6 / 17)	1.16 (.55 - 2.48)	31.51% (23 / 73)
Sore Throat	27.59% (16 / 58)	22.22% (4 / 18)	.81 (.31 - 2.10)	26.32% (20 / 76)
Nausea	48.28% (28 / 58)	41.18% (7 / 17)	.85 (.46 - 1.60)	46.67% (35 / 75)
Vomitting	31.03% (18 / 58)	27.78% (5 / 18)	.90 (.39 - 2.07)	30.26% (23 / 76)
Diarrhea	50% (29 / 58)	35.29% (6 / 17)	.71 (.35 - 1.41)	44% (33 / 75)
Abdominal Pain	31.03% (18 / 58)	5.88% (1 / 17)	.19 (.03 - 1.32)	25.33% (19 / 75)
Hypogeusia	37.04% (20 / 54)	33.33% (5 / 15)	.90 (.41 - 1.99)	36.23% (25 / 69)
Anosmia	31.37% (16 / 51)	33.33% (5 / 15)	1.06 (.47 - 2.42)	31.82% (21 / 66)
All Cause Mortality	3.75% (3 / 80)	27.27% (9 / 33)	7.27 *** (2.10 - 25.19) [p = .0002]	10.62% (12 / 113)

Unless otherwise noted, relative risks were not statistically significant. Moderate (clinical score 1–3) and severe (clinical score 4–5) disease status were assigned as described in Methods. Percentages of sub-group (moderate or severe) are shown for each category with respective counts in parenthesis. Average age was calculated with accompanying sample standard deviation. Ethnicity and BMI were extracted from most recent electronic medical record (EMR) data. Select COVID-19 risk factors were scored by a clinical infectious disease physician. Presenting symptoms were recorded through direct interview with patient or surrogate or retrospective EMR review.

nature research

Corresponding author(s): Akiko Iwasaki

Last updated by author(s): Jun 23, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	EPIC EHR software (retrospective EMR review and clinical data aggregation) and REDCap 9.3.6 (clinical data aggregation).
Data analysis	GraphPad PRISM version 8.0.2 (statistics/graphics), R 3.4.3 (graphs/statistics), JMP15 (graphs), ggplot2, caret, tidyverse, ggpubr, Igraph, mlbench, and ggstatsplot, FlowJo software version 10.6 software (Tree Star).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data generated during the current study will be available before publication in a public repository.
Accession code number: SDY1655

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to calculate the sample size. Sample size was determined based on the number of patients admitted to Yale-New Haven Hospital (YNHH) between March 18th and May 5th that were enrolled and consented with th current study. This study enrolled 135 patients admitted to the Yale New Haven Health care network under IRB and HIC approved protocol #2000027690. Patients were identified though screening of EMR records for potential enrollment. Informed consent was obtained by trained staff and sample collection commenced immediately upon study enrollment. Clinical specimens were collected approximately every 4 days where an individual's clinical status permitted, and was continued until patient discharge or expiration.
Data exclusions	135 COVID-19 patients were enrolled on this study however 22 were excluded. Those included: Pregnant women and patients on active chemotherapy. Specifically, cytokine ELISAs from two individuals were excluded from analysis due to poor sample quality. Measurements from these individuals were outliers (beyond 1.5x the interquartile range) in more than half of the cytokines measured. This strongly suggested that a technical error occurred during these two experiments.Finally, for each individual boxplot, line graph, or linear regression, unique values that fell into the top or bottom 1% were excluded. Duplicate values within this range were not excluded. This applies only to unique values, such that two identical measurements falling into this range will remain in the analysis. We chose this very conservative method of exclusion in order to most faithfully represent the heterogeneity of our data, without allowing for extreme outliers to obscure our analyses. This is particularly true in situations in which we subset the data further by time intervals; with a smaller n in each time interval, extreme outliers disproportionately skew the mean/median at this point. Finally for the health donors group, asymptomatic or pre-symptomatic healthcare workers were excluded (when positive for SARS-CoV2 q-RT-PCR or serology).
Replication	The findings were not replicated - longitudinal analyses from human individuals.
Randomization	Patients were stratified by disease severity (moderate and severe) based on based on oxygen levels and intensive care unit (ICU) requirement. Moderate disease status (Clinical Score 1, 2 and 3) was defined as: (1) SARS-CoV-2 infection requiring hospitalization without supplemental oxygen, (2) infection requiring non-invasive supplemental oxygen (<3 L / min, sufficient to maintain greater than 92% SpO2), (3) infection requiring non-invasive supplemental oxygen (> 3L supplemental oxygen to maintain SpO2 > 92%, or, required > 2L supplemental oxygen to maintain SpO2 > 92% and had a high sensitivity C-reactive protein (CRP) > 70) and received tocilizumab. Severe disease status (Clinical score 4 and 5) was defined as infection meeting all criteria for clinical score 3 while also requiring admission to the YNHH Intensive Care Unit (ICU) and > 6L supplemental oxygen to maintain SpO2 > 92% (4); or infection requiring invasive mechanical ventilation / extracorporeal membrane oxygenation (ECMO) in addition to glucocorticoid / vasopressor administration (5). Clinical score 6 was assigned for deceased patients.
Blinding	At the time of sample acquisition and processing, scientists were completely unaware of the patients' conditions. Blood acquisition is performed and recorded by a separate team. Information of patients' conditions are not available until after processing and analysing raw data by flow cytometry and ELISA. A clinical team, separate from the experimental team, performs chart review to determine patients' relevant statistics. Cytokines and facs analyses were blinded. Patients clinical information and clinical scores coding were only revealed after data collection.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies used in this study are against human proteins. BB515 anti-hHLA-DR (G46-6) (1:400) (BD Biosciences), BV785 anti-hCD16 (3G8) (1:100) (BioLegend), PE-Cy7 anti-hCD14 (HCD14) (1:300) (BioLegend), BV605 anti-hCD3 (UCHT1) (1:300) (BioLegend), BV711 anti-hCD19 (SJ25C1) (1:300) (BD Biosciences), AlexaFluor647 anti-hCD1c (L161) (1:150) (BioLegend), Biotin anti-hCD141 (M80) (1:150) (BioLegend), PE-Dazzle594 anti-hCD56 (HCD56) (1:300) (BioLegend), PE anti-hCD304 (12C2) (1:300) (BioLegend), APCFire750 anti-hCD11b (ICRF44) (1:100) (BioLegend), PerCP/Cy5.5 anti-hCD66b (G10F5) (1:200) (BD Biosciences), BV785 anti-hCD4 (SK3) (1:200) (BioLegend), APCFire750 or PE-Cy7 or BV711 anti-hCD8 (SK1) (1:200) (BioLegend), BV421 anti-hCCR7 (G043H7) (1:50) (BioLegend), AlexaFluor 700 anti-hCD45RA (HI100) (1:200) (BD Biosciences), PE anti-hPD1 (EH12.2H7) (1:200) (BioLegend), APC anti-hTIM3 (F38-2E2) (1:50) (BioLegend), BV711 anti-hCD38 (HIT2) (1:200) (BioLegend), BB700 anti-hCXCR5 (RF8B2) (1:50) (BD Biosciences), PE-Cy7 anti-hCD127 (HIL-7R-M21) (1:50) (BioLegend), PE-CF594 anti-hCD25 (BC96) (1:200) (BD Biosciences), BV711 anti-hCD127 (HIL-7R-M21) (1:50) (BD Biosciences), BV421 anti-hIL17a (N49-653) (1:100) (BD Biosciences), AlexaFluor 700 anti-hTNFa (MAb11) (1:100) (BioLegend), PE or APC/Fire750 anti-hIFNy (4S.B3) (1:60) (BioLegend), FITC anti-hGranzymeB (GB11) (1:200) (BioLegend), AlexaFluor 647 anti-hIL-4 (8D4-8) (1:100) (BioLegend), BB700 anti-hCD183/CXCR3 (1C6/CXCR3) (1:100) (BD Biosciences), PE-Cy7 anti-hIL-6 (MQ2-13A5) (1:50) (BioLegend), PE anti-hIL-2 (5344.111) (1:50) (BD Biosciences), BV785 anti-hCD19 (SJ25C1) (1:300) (BioLegend), BV421 anti-hCD138 (MI15) (1:300) (BioLegend), AlexaFluor700 anti-hCD20 (2H7) (1:200) (BioLegend), AlexaFluor 647 anti-hCD27 (M-T271) (1:350) (BioLegend), PE/Dazzle594 anti-hIgD (IA6-2) (1:400) (BioLegend), PE-Cy7 anti-hCD86 (IT2.2) (1:100) (BioLegend), APC/Fire750 anti-hIgM (MHM-88) (1:250) (BioLegend), BV605 anti-hCD24 (ML5) (1:200) (BioLegend), BV421 anti-hCD10 (HI10a) (1:200) (BioLegend), BV421 anti-CDh15 (SSEA-1) (1:200) (BioLegend), AlexaFluor 700 Streptavidin (1:300) (ThermoFisher), BV605 Streptavidin (1:300) (BioLegend).

Validation

All antibodies used in this study are commercially available, and all have been validated by the manufacturers and used by other publications. Likewise, we titrated these antibodies according to our own our staining conditions. The following were validated in the following species: BB515 anti-hHLA-DR (G46-6) (BD Biosciences) (Human, Rhesus, Cynomolgus, Baboon), BV785 anti-hCD16 (3G8) (BioLegend) (Human, African Green, Baboon, Capuchin Monkey, Chimpanzee, Cynomolgus, Marmoset, Pigtailed Macaque, Rhesus, Sooty Mangabey, Squirrel Monkey), PE-Cy7 anti-hCD14 (HCD14) (BioLegend) (Human), BV605 anti-hCD3 (UCHT1) (BioLegend) (Human, Chimpanzee), BV711 anti-hCD19 (SJ25C1) (BD Biosciences) (Human), AlexaFluor647 anti-hCD1c (L161) (BioLegend) (Human, African Green, Baboon, Cynomolgus, Rhesus), Biotin anti-hCD141 (M80) (BioLegend) (Human, African Green, Baboon), PE-Dazzle594 anti-hCD56 (HCD56) (BioLegend) (Human, African Green, Baboon, Cynomolgus, Rhesus), PE anti-hCD304 (12C2) (BioLegend) (Human), APCFire750 anti-hCD11b (ICRF44) (BioLegend) (Human, African Green, Baboon, Chimpanzee, Common Marmoset, Cynomolgus, Rhesus, Swine), PerCP/Cy5.5 anti-hCD66b (G10F5) (BD Biosciences) (Human), BV785 anti-hCD4 (SK3) (BioLegend) (Human), APCFire750 or PE-Cy7 or BV711 anti-hCD8 (SK1) (BioLegend) (Human, Cross-Reactivity: African Green, Chimpanzee, Cynomolgus, Pigtailed Macaque, Rhesus, Sooty Mangabey), BV421 anti-hCCR7 (G043H7) (BioLegend) (Human, African Green, Baboon, Cynomolgus, Rhesus), AlexaFluor 700 anti-hCD45RA (HI100) (BD Biosciences) (Human), PE anti-hPD1 (EH12.2H7) (BioLegend) (Human, African Green, Baboon, Chimpanzee, Common Marmoset, Cynomolgus, Rhesus, Squirrel Monkey), APC anti-hTIM3 (F38-2E2) (BioLegend) (Human), BV711 anti-hCD38 (HIT2) (BioLegend) (Human, Chimpanzee, Horse), BB700 anti-hCXCR5 (RF8B2) (BD Biosciences) (Human), PE-Cy7 anti-hCD127 (HIL-7R-M21) (BioLegend) (Human), PE-CF594 anti-hCD25 (BC96) (BD Biosciences) (Human, Rhesus, Cynomolgus, Baboon), BV711 anti-hCD127 (HIL-7R-M21) (BD Biosciences) (Human), BV421 anti-hIL-17a (N49-653) (BD Biosciences) (Human), AlexaFluor 700 anti-hTNFa (MAb11) (BioLegend) (Human, Cat, Cross-Reactivity: Chimpanzee, Baboon, Cynomolgus, Rhesus, Pigtailed Macaque, Sooty Mangabey, Swine), PE or APC/Fire750 anti-hIFNy (4S.B3) (BioLegend) (Human, Cross-Reactivity: Chimpanzee, Baboon, Cynomolgus, Rhesus), FITC anti-hGranzymeB (GB11) (BioLegend) (Human, Mouse, Cross-Reactivity: Rat), AlexaFluor 647 anti-hIL-4 (8D4-8) (BioLegend) (Human, Cross-Reactivity: Chimpanzee, Baboon, Cynomolgus, Rhesus), BB700 anti-hCD183/CXCR3 (1C6/CXCR3) (BD Biosciences) (Human, Rhesus, Cynomolgus, Baboon), PE-Cy7 anti-IL-6 (MQ2-13A5) (BioLegend) (Human), PE anti-hIL-2 (5344.111) (BD Biosciences) (Human), BV785 anti-hCD19 (SJ25C1) (BioLegend) (Human), BV421 anti-hCD138 (MI15) (BioLegend) (Human), AlexaFluor700 anti-hCD20 (2H7) (BioLegend) (Human, Baboon, Capuchin Monkey, Chimpanzee, Cynomolgus, Pigtailed Macaque, Rhesus, Squirrel Monkey), AlexaFluor 647 anti-hCD27 (M-T271) (BioLegend) (Human, Cross-Reacitivity: Baboon, Cynomolgus, Rhesus), PE/Dazzle594 anti-hIgD (IA6-2) (BioLegend) (Human), PE-Cy7 anti-hCD86 (IT2.2) (BioLegend) (Human, African Green, Baboon, Capuchin Monkey, Common Marmoset, Cotton-topped Tamarin, Chimpanzee, Cynomolgus, Rhesus), APC/Fire750 anti-hIgM (MHM-88) (BioLegend) (Human, African Green, Baboon, Cynomolgus, Rhesus), BV605 anti-hCD24 (ML5) (BioLegend) (Human, Cross-Reactivity: Chimpanzee), BV421 anti-hCD10 (HI10a) (BioLegend) (Human, African Green, Baboon, Capuchin monkey, Chimpanzee, Cynomolgus, Rhesus), BV421 anti-hCD15 (SSEA-1) (BioLegend) (Human), AlexaFluor 700 Streptavidin (1:300) (ThermoFisher), BV605 Streptavidin (1:300) (BioLegend).

Human research participants

Policy information about studies involving human research participants

Population characteristics

Cohort characteristics: age (62.96 ± 17.0), sex (Male 46.02% / Females 53.98% , Ethnicity (American Indian -Alaskan Native 0%/ Asian (0.88%) / Black -African American (29.2%)/ Native Hawaiian-Pacific Islander(0%)/ White (53.98%)/ Hispanic (12.39%). Full demographic data is included in Extended data table 1.

Recruitment

Patients admitted to the Yale New Haven Hospital (YNHH) between the 18th of March through the 27th of May 2020, were recruited to the Yale IMPACT study (Implementing Medical and Public Health Action Against Coronavirus CT) after testing positive for SARS-CoV2 by qRT-PCR. (serology was further confirmed for all patients enrolled). Patients were identified though screening of EMR records for potential enrollment with no self selection. Informed consent was obtained by trained staff and sample collection commenced immediately upon study enrollment. Clinical specimens were collected approximately every 4 days where an individual's clinical status permitted, and was continued until patient discharge or expiration.

Ethics oversight

Yale Human Research Protection Program Institutional Review Boards. Informed consents were obtained from all enrolled patients and healthcare workers. • Our research protocol was reviewed and approved by the Yale School of Medicine IRB and HIC (#2000027690). Informed consent was obtained by trained staff and records maintained in our research database for the duration of our study. There were no minors included on this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Freshly isolated PBMCs were stained for live and dead markers, blocked with Human TruStan FcX , stained for surface markers and then fixed with PFA 4%. For intracellular cytokine staining following stimulation , cells were surface stained, washed and fixed in 4% PFA. After permeabilization with 1X Permeabilization Buffer cells were stained for intracellular cytokines analysis.
Instrument	Cells were acquired on an Attune NXT (ThermoFisher).
Software	Data were analysed using FlowJo software version 10.6 software (Tree Star). n
Cell population abundance	Cell population abundance: Cells populations were reported in various formats including as a number or concentration of the patient's blood sample (x106cells/mL), as a proportion of live, single PBMC (% of Live), or as a proportion of a parent gate (% of CD4 T cells, % of Monocytes, etc.). The full gating path for clarification is included in the extended figures.
Gating strategy	SSC-A and FSC-A parameters were used to select leukocytes from isolated PBMCs. Live and dead cells were defined based on aqua staining. Singlets were separated based on SSC/ FSC parameters. Leukocytes were gated based on to identify lymphocytes (CD3/CD4/CD8/CD19/CD56 markers), granulocytes (CD16,CD14, HLA-DR markers) and pDCs, and cDCs (CD304, CD1c, CD141). TCR-activated T cells, Terminally-differentiated T cells, and additional subsets.were defined using HLA-DR, CD38, CCR7,CD127, PD1, TIM-3, CXCR5, CD45RA, CD25. Intracellular T cell gating strategy to identify CD4 and/or CD8 T cells secreting TNFa, IFN-y, IL-6, IL-2, GranzymeB, IL-4, and/or IL-17 were defined using the specif markers: CD3, CD4, CD8, TNF, IFN, IL-6, IL-2, IL-4, IL-17 and granzyme B.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

A universal trade-off between growth and lag in fluctuating environments

https://doi.org/10.1038/s41586-020-2505-4

Received: 4 April 2018

Accepted: 21 April 2020

Markus Basan[1,2,6] ✉, Tomoya Honda[3,6], Dimitris Christodoulou[2], Manuel Hörl[2], Yu-Fang Chang[1], Emanuele Leoncini[1], Avik Mukherjee[1], Hiroyuki Okano[4], Brian R. Taylor[4], Josh M. Silverman[5], Carlos Sanchez[1], James R. Williamson[5], Johan Paulsson[1], Terence Hwa[3,4] ✉ & Uwe Sauer[2] ✉

The rate of cell growth is crucial for bacterial fitness and drives the allocation of bacterial resources, affecting, for example, the expression levels of proteins dedicated to metabolism and biosynthesis[1,2]. It is unclear, however, what ultimately determines growth rates in different environmental conditions. Moreover, increasing evidence suggests that other objectives are also important[3–7], such as the rate of physiological adaptation to changing environments[8,9]. A common challenge for cells is that these objectives cannot be independently optimized, and maximizing one often reduces another. Many such trade-offs have indeed been hypothesized on the basis of qualitative correlative studies[8–11]. Here we report a trade-off between steady-state growth rate and physiological adaptability in *Escherichia coli*, observed when a growing culture is abruptly shifted from a preferred carbon source such as glucose to fermentation products such as acetate. These metabolic transitions, common for enteric bacteria, are often accompanied by multi-hour lags before growth resumes. Metabolomic analysis reveals that long lags result from the depletion of key metabolites that follows the sudden reversal in the central carbon flux owing to the imposed nutrient shifts. A model of sequential flux limitation not only explains the observed trade-off between growth and adaptability, but also allows quantitative predictions regarding the universal occurrence of such tradeoffs, based on the opposing enzyme requirements of glycolysis versus gluconeogenesis. We validate these predictions experimentally for many different nutrient shifts in *E. coli*, as well as for other respiro-fermentative microorganisms, including *Bacillus subtilis* and *Saccharomyces cerevisiae*.

To study the interrelationship between the rate of cell growth and the rate of physiological adaptation (the latter being characterized by the inverse of the 'lag time' defined in Fig. 1a), we shifted wild-type *E. coli* (Supplementary Table 1) between two minimal media, each containing a single carbon source. Defined postshift conditions and very rapid environmental changes were implemented as 'complete shifts' that ensured that no preshift carbon source was available to cells in the postshift medium (Fig. 1b). We first investigated shifts from different glycolytic carbon sources to acetate, a gluconeogenic carbon source that requires fluxes through glycolysis to reverse direction. Because acetate is the primary fermentation product of many bacteria, including *E. coli*, it is naturally available to these bacteria upon exhaustion of the primary carbon source.

We quantified these shifts by lag time, defined as the integrated time lost during the adaptation to new conditions compared with an immediate response (Fig. 1a). We found that the shifts produced extended lags of up to 10 h (Fig. 1c, circles), much longer than the doubling times in preshift and postshift media (less than 2 h), and often included periods without detectable biomass production that lasted several hours (Extended Data Fig. 1a). A notable correlation emerged between the growth rate in the preshift medium and the lag time (Fig. 1c, circles): fast-growing cells took a long time to adjust to the new medium, whereas slow-growing cells resumed growth much more quickly. The same relation was obtained when the preshift growth was varied by titrating the uptake rates of lactose as an example of a glycolytic carbon source (Fig. 1c, squares), suggesting that the relation between preshift growth and lag time depends on the carbon influx rate rather than on the specifics of the preshift carbon sources. A similar pattern was found for population growth dynamics with chemostat-controlled growth rates[12]. The data in Fig. 1c show that lag times (T_{lag}) increased with increasing preshift growth rate (λ_{pre}), with an apparent divergence at a critical growth rate, λ_0. Indeed, replotting the data of Fig. 1c reveals an approximately linear relation (Fig. 1d, purple

---

[1]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [2]Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland. [3]Section of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, CA, USA. [4]Department of Physics, University of California at San Diego, La Jolla, CA, USA. [5]Department of Integrative Structural and Computational Biology, and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA. [6]These authors contributed equally: Markus Basan, Tomoya Honda. ✉e-mail: markus@hms.harvard.edu; thwa@ucsd.edu; sauer@imsb.biol.ethz.ch

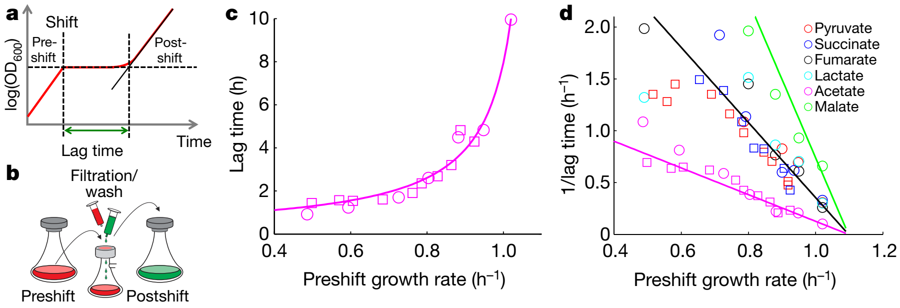


Fig. 1 | Phenomenological characterization of lag phase. a, Illustration of a typical growth curve. The lag time is defined as the time lost during transition to new conditions (from preshift to postshift) as compared with an instantaneous switch to final steady-state growth. OD$_{600}$, optical density at 600 nm. **b**, Illustration of our medium-transfer protocol. **c**, Circles show lag times of the wild type after shifts from different glycolytic carbon sources to acetate minimal medium. Squares show lag times resulting when preshift growth is instead varied by titrating the uptake rates of lactose as an example of a glycolytic carbon source (using *E. coli* strain NQ381, which has a titratable lactose-uptake system). The preshift glycolytic carbon sources—ordered from fast growth rates to slow growth rates—are glucose-6-phosphate, glucose, mannitol, maltose, glycerol, galactose and mannose, which are all readily metabolized by wild-type *E. coli*, yet result in very different growth rates. The solid line represents the empirical relation given by equation (1). **d**, Inverse lag times for shifts from different glycolytic to gluconeogenic carbon sources, plotted against preshift growth rates. Colours indicate shifts to the postshift carbon sources shown in the inset; different circles of the same colour indicate different preshift carbon sources, and squares indicate the use of titratable lactose uptake in preshift. Lines show nonlinear least-squares mean fits of equation (1) to lag-time data as a function of preshift growth rates for the shifts to acetate (magenta line) and to succinate and pyruvate (black line) from our batch culture experiments (Supplementary Table 2), assuming a λ_C of approximately 1.1 h^{-1}. For the shift to malate, we performed an additional fit, again assuming a λ_C of approximately 1.1 h^{-1} (green line). Nonlinear least-squares mean fits of equation (1) to individual shifts are shown in Extended Data Fig. 2 and the resulting 95% confidence intervals of parameters are as follows: acetate, $\lambda_C = (1.10 \pm 0.01)$ h^{-1}, $\alpha = 0.78 \pm 0.10$, $n = 17$; pyruvate, $\lambda_C = (1.12 \pm 0.03)$ h^{-1}, $\alpha = 0.33 \pm 0.07$, $n = 17$; succinate, $\lambda_C = (1.13 \pm 0.04)$ h^{-1}, $\alpha = 0.33 \pm 0.09$, $n = 14$; fumarate, $\lambda_C = (1.08 \pm 0.02)$ h^{-1}, $\alpha = 0.23 \pm 0.07$, $n = 5$; lactate, $\lambda_C = (1.09 \pm 0.05)$ h^{-1}, $\alpha = 0.22 \pm 0.15$, $n = 5$; malate, $\lambda_C = (1.17 \pm 0.09)$ h^{-1}, $\alpha = 0.22 \pm 0.11$, $n = 5$. The mean critical growth rate and standard deviation resulting from the individual fits are given by $\lambda_C = (1.11 \pm 0.03)$ h^{-1}.

symbols and line) between the inverse lag time, $1/T_{\mathrm{lag}}$, a measure of adaptability, and λ_{pre}, that is:

$$\frac{1}{T_{\mathrm{lag}}} = \alpha \cdot (\lambda_0 - \lambda_{\mathrm{pre}}), \qquad (1)$$

in which α is a dimensionless proportionality constant.

To test the generality of this relation, we analysed lag times in 144 transitions (Supplementary Tables 2, 3), finding long lag times for shifts from six glycolytic to six gluconeogenic carbon sources (Extended Data Fig. 2a–f). Notably, all of these shifts exhibited similar linear relations between the preshift growth rate and the inverse lag time, but with different proportionality constants, α, for different postshift carbon sources, all with the same critical growth rate, λ_0, of approximately 1.1 doublings per hour (Fig. 1d and Extended Data Fig. 2). Some degree of correlation also exists between the lag time and postshift growth rates (Extended Data Fig. 2g), as observed previously[13], but the pattern is much weaker compared with those seen in Fig. 1c, d. We also examined several classic diauxic shifts, where both carbon sources were present in preshift, and found the lag times in most cases to be very similar to those for the complete shifts that we study here (Extended Data Fig. 1b–d).

To investigate the origin of the extended lag time in our shifts, we first tested whether dormant and heterogeneous subpopulations may play a part. Using two complementary methods (Supplementary Note 1 and Extended Data Figs. 3, 4), we quantified cell-to-cell variability following the shift from glucose to acetate. The results revealed some heterogeneity in lag times, but no distinct subpopulations: none of the cells resumed growth immediately after the shift, and virtually all cells resumed growth shortly after the average lag time.

To determine whether the observed correlation between lag time and preshift growth is due to a limitation in central metabolism (referred to as a 'metabolic limitation'), we quantified metabolite pools throughout the lag phase of the glucose-to-acetate transition (Fig. 2a). By comparing the dynamics of metabolite pools and fluxes with steady-state levels during exponential growth on glucose and acetate, we can infer metabolic bottlenecks. Over the course of the lag phase, the concentrations of different metabolites increased in a sequential manner (Fig. 2b) that matched their position in gluconeogenesis: metabolites in the tricarboxylic acid (TCA) cycle (citrate and malate) started to accumulate at 1–2 h into the lag phase, and also overshot their postshift steady-state values (Fig. 2b, dashed black line) by several-fold once growth resumed at approximately 4 h after shift (Fig. 2a). The levels of metabolites in upper glycolysis increased even later (Fig. 2b and Extended Data Fig. 5a). Notably, the increase in the latter coincided with the time of growth resumption (Fig. 2a). In particular, the pool of the key regulatory metabolite fructose-1,6-bisphosphate (FBP) plunged rapidly by 200-fold within 30 min of the shift and remained well below its postshift steady-state level until 30 min before growth resumption (Extended Data Fig. 5c). This finding is not compatible with the mechanism recently proposed to underlie lag phases to gluconeogenesis based on a postulated high FBP pool in the majority of the cell population during lag phase[12].

Estimating the fluxes by multiplying measured metabolite concentrations and the turnover rates derived from ^{13}C-labelling dynamics, we observed a sequential pattern that followed their position in gluconeogenesis (Fig. 2c). TCA cycle metabolites quickly became fully ^{13}C-labelled. By contrast, a gluconeogenic flux to upper glycolysis was hardly detectable even 30 min after the shift, and was still below 1% of the steady-state level 1.5 hours after shift.

The observed metabolic dynamics suggest that gluconeogenic flux limits the biosynthesis of biomass components derived from intermediates in upper glycolysis. In particular, metabolites such as erythrose-4-phosphate and ribose-5-phosphate—which branch off from upper glycolysis and are required for the biosynthesis of specific amino acids and nucleotides—may limit biomass production. Because biomass synthesis requires fixed stoichiometric ratios of building blocks, metabolites in the TCA cycle and lower glycolysis accumulate far beyond their steady-state concentrations (Fig. 2b), as they cannot be incorporated into biomass in the absence of sufficient metabolites from upper glycolysis. In accordance with this hypothesis, we found the absolute concentrations of key metabolites in upper glycolysis (for example, F6P) to be small compared with the affinity constants of the key enzymes required to produce erythrose-4-phosphate and ribose-5-phosphate (Supplementary Table 4).

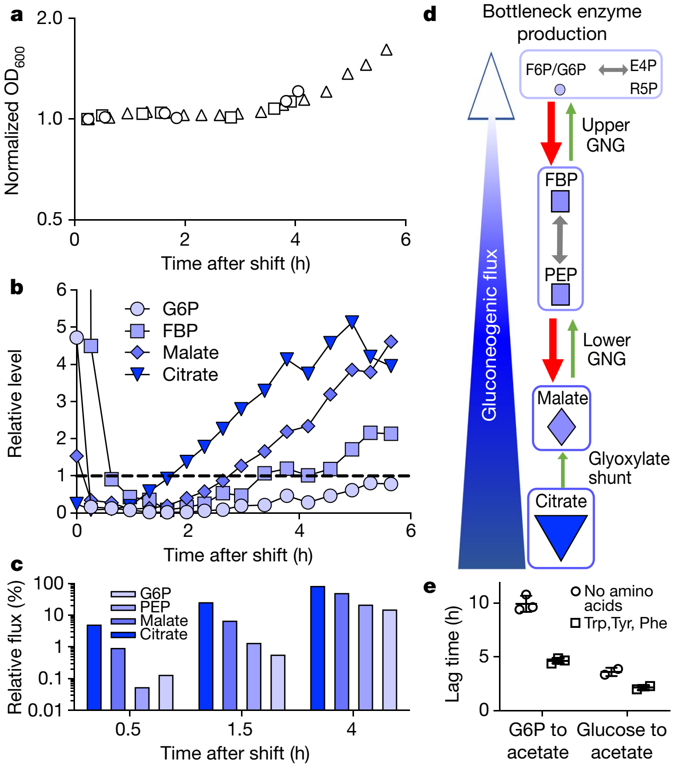


Fig. 2 | Metabolic characterization of lag phase during shifts to acetate. **a**, Normalized cell density during lag phase following three shifts from glucose to acetate, used for metabolite measurements (triangles) and for flux measurements (squares and circles). **b**, Temporal profiles of metabolites—glucose-6-phosphate (G6P), FBP, malate and citrate—throughout lag phase following a shift from glucose to acetate, normalized by their respective values in postshift medium during exponential steady-state growth (dashed line). Steady-state metabolite concentrations during exponential growth were measured in separate experiments by taking three metabolite measurements throughout the exponential growth curve from each of two biological repeats. The metabolite concentrations during the lag phase were then normalized by these steady-state concentrations. Time zero values are measured preshift levels. For FBP, this value (approximately 157) falls outside the scale. **c**, Fluxes to different metabolites (**b**) at three time points during the lag phase from glucose to acetate, as a percentage of steady-state flux during growth on acetate (measured in separate steady-state experiments for two biological repeats). **d**, Illustration of glycolysis/gluconeogenesis. The large fading blue arrow indicates the directionality of gluconeogenesis and illustrates the decrease in normalized fluxes and metabolite pools. Green arrows indicate irreversible gluconeogenic reactions catalysed by gluconeogenic enzymes (GNGs); red arrows indicate the residual activity of glycolytic enzymes acting in the opposite direction. Erythrose-4-phosphate (E4P) and ribose-5-phosphate (R5P) are derived from fructose-6-phosphate (F6P)/G6P and are required for the biosynthesis of specific amino acids and nucleotides. PEP, phosphoenolpyruvate. **e**, The addition of three non-degradable amino acids derived from upper glycolysis—tyrosine (Tyr), tryptophan (Trp) and phenylalanine (Phe)—to the postshift growth medium substantially reduces lag times in shifts to acetate from preshift growth on glucose and on G6P.

After the shift to acetate, gluconeogenic flux is essential for biomass production and enzyme synthesis. Although many glycolytic enzymes can operate reversibly and can thereby also catalyse gluconeogenesis, several glycolytic reactions are thermodynamically strongly favoured in the glycolytic direction, such that they can be considered effectively irreversible. As illustrated in Fig. 2d, in a simplified picture of central metabolism, gluconeogenesis can be considered as a linear pathway consisting of 'lower gluconeogenenic' reactions (catalysed by phosphoenolpyruvate carboxykinase, Pck; malate dehydrogenases, MaeA and MaeB; and phosphoenolpyruvate synthetase, Pps) and 'upper gluconeogenic' reactions (catalysed primarily by the essential enzyme fructose-1,6-bisphosphatase, Fbp). These dedicated gluconeogenic enzymes are required for gluconeogenesis, but many of them are expressed at low levels during preshift growth and immediately after the shift when compared with their abundances in the postshift steady state (Extended Data Fig. 6), presumably because the activities of the gluconeogenic enzymes can lead to substantial futile cycling that dissipates energy. Consistent with the observed increase in lag time with higher preshift growth rates (Fig. 1c), the abundances of the lower gluconeogenic enzymes (quantified previously through proteomics[3]) decrease with higher preshift growth rates (Fig. 3a).

Quantitative proteomics measurements showed that the abundances of gluconeogenic enzymes increased very gradually, coinciding with exit from the lag phase (Extended Data Fig. 6). During the lag phase, formation of these lower gluconeogenic enzymes requires precursors (for example, specific amino acids), whose synthesis rate is in turn limited by the gluconeogenic flux. Hence, right after the shift, the cell is trapped in a state in which a bottleneck in gluconeogenic flux limits the synthesis of amino acids and hence the production of enzymes needed to alleviate the bottleneck (Extended Data Fig. 7a). Indeed, reducing the requirements of metabolites resulting from gluconeogenic flux, such as erythrose-6-phosphate, by adding the three aromatic amino acids derived from it (tryptophan, phenylalanine and tyrosine) to the postshift medium (Fig. 2e) reduced the lag time by roughly 50%, even though individually these amino acids do not support growth[14].

For rapid adaptations dominated by simple catabolic bottlenecks, a kinetic model of growth adaptation based on the dynamic reallocation of proteomic resources has been shown to give quantitatively accurate descriptions of adaptation dynamics[15]. However, for the very long lag phases studied here, severe internal metabolic bottlenecks are involved owing to the reversal of central carbon fluxes. Guided by the metabolomic and proteomic data (Fig. 2), we constructed a minimalistic mathematical model. We assumed that the gluconeogenic flux is the bottleneck for the amino-acid synthesis required for de novo production of gluconeogenic enzymes during the lag phase (illustrated in Extended Data Fig. 7a and resulting in the equation therein). As illustrated in Extended Data Fig. 7b and explained in Supplementary Note 2, the gluconeogenic flux is determined by the scaling of metabolite concentrations at lower and upper gluconeogenesis, which are in turn determined by the levels of lower gluconeogenic enzymes, resulting in the equations in Extended Data Fig. 7b. Solving the resulting differential equation, we arrive at a simple expression for the inverse lag time:

$$\frac{1}{T_{\mathrm{lag}}} \propto \phi^{\mathrm{pre}}_{\mathrm{GNG,\,lower}} \tag{2}$$

in which $\phi^{\mathrm{pre}}_{\mathrm{GNG,\,lower}}$ denotes the preshift abundance of lower gluconeogenic enzymes that provide the initial condition. The abundances of these enzymes rise throughout the lag phase (Extended Data Fig. 6), and their abundances in preshift conditions[14] are well-described by a linear decrease with increasing preshift growth rate, λ_{pre}, that is:

$$\phi^{\mathrm{pre}}_{\mathrm{GNG,\,lower}} \propto (\lambda_C - \lambda_{\mathrm{pre}})$$

in which $\phi^{\mathrm{pre}}_{\mathrm{GNG,\,lower}}$ is vanishing at a characteristic growth rate, λ_C, of approximately 1.1 h^{-1} (Fig. 3a, lines). This resembles the linear cyclic AMP (cAMP)-mediated increase in catabolic protein abundances for carbon-limited growth[14]. Inserting this growth-rate dependence into equation (2), we obtain $1/T_{\mathrm{lag}} \propto (\lambda_C - \lambda_{\mathrm{pre}})$, which is identical to the empirical relation equation (1), with the same critical growth rate λ_0 of roughly 1.1 h^{-1}. Thus, our model successfully recapitulates the observed growth-rate/lag-time relations (Fig. 1d) up to an overall scaling factor, α (equation (1)).

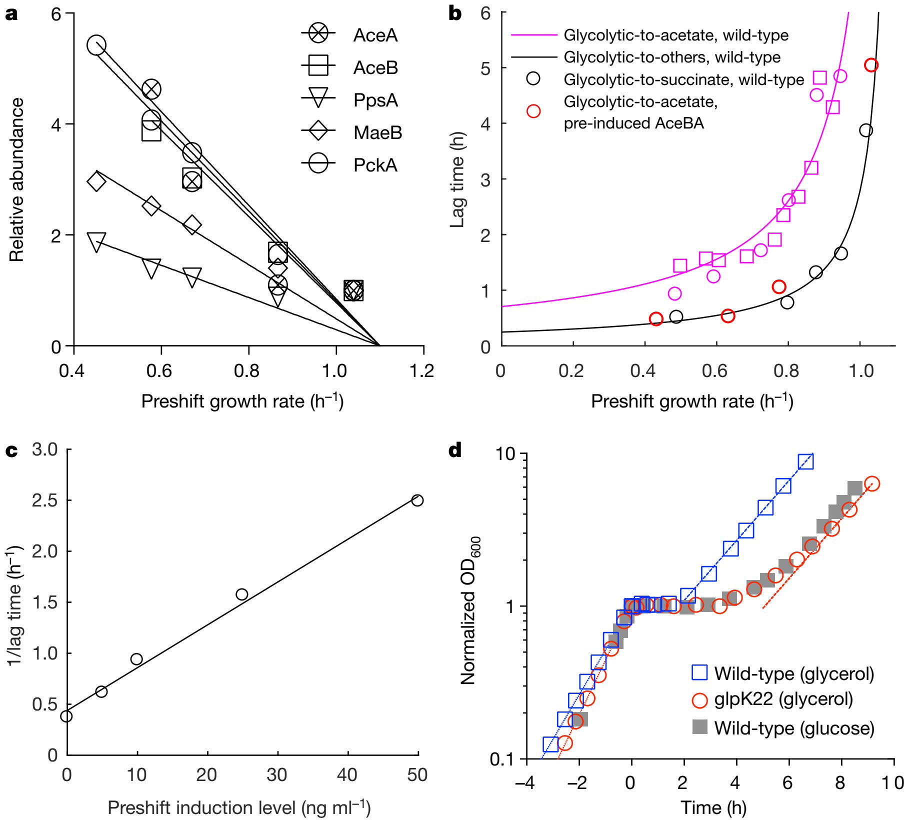


Fig. 3 | Tests of model predictions. a, Relative abundance of gluconeogenic enzymes at different growth rates during steady-state exponential growth in glycolytic conditions; data from ref. [3]. Enzymes are isocitrate lyase (AceA), malate synthase A (AceB), phosphoenolpyruvate synthetase (PpsA), malate dehydrogenase (MaeB) and phosphoenolpyruvate carboxykinase (PckA). The lines are linear fits assuming a characteristic growth rate, λ_C, at which lower gluconeogenic enzymes are not expressed anymore, given by $\lambda_C \approx 1.1\ h^{-1}$, identical to the critical growth rate at which lag times diverge, $\lambda_0 \approx 1.1\ h^{-1}$ (determined in Fig. 1c). **b**, Lag times during shifts from various carbon sources to gluconeogenic carbon sources. Magenta lines and symbols represent shifts to acetate for wild-type cells (data shown in Fig. 1c, d). Bold red symbols represent reduced lag times for shifts to acetate for a strain with preshift expression of enzymes of the glyoxylate shunt, AceBA. Those data fall on the black line, which is the trendline of lag times for shifts of the wild type to other gluconeogenic carbon sources (Fig. 1d). As an example, the black symbols represent shifts to succinate. **c**, Inverse lag times for shifts from glucose to pyruvate, plotted against different preshift PpsA induction levels, for a strain harbouring titratable PpsA expression. **d**, Growth of strain NQ898, which contains the glycerol-uptake mutation *glpK22* (red circles)[17], is faster than that of the wild-type strain NCM3722 in preshift glycerol medium (0.82 h^{-1} versus 0.68 h^{-1}), but the lag time (as defined in Fig. 1b) upon abrupt shift to acetate at time $t = 0$ is substantially longer (5.1 h versus 1.9 h). For comparison, the transition of the wild-type strain grown in preshift glucose medium (0.87 h^{-1}) to acetate is shown in grey. The dashed lines indicate the steady-state growth rates of the two strains in acetate, both about 0.45 h^{-1}.

Lag times for most postshift carbon sources collapse on the same curve (Fig. 1d and Extended Data Fig. 2, black lines). However, shifts to acetate are described by a different scaling factor α (magenta symbols and line), and shifts to malate show a milder deviation (green circles and line). A possible explanation for the altered acetate line is that only growth on acetate requires the glyoxylate shunt in addition to other gluconeogenic enzymes. If true, then pre-expressing enzymes of the glyoxylate shunt (AceB and AceA) should eliminate this additional bottleneck and revert the relation between lag time and growth rate to that observed for shifts to most other TCA cycle substrates (Fig. 1d and Extended Data Fig. 2, black line). Indeed, preshift expression of the glyoxylate bypass reduced the lag times for various shifts to acetate (compare red circles and magenta curve in Fig. 3b), such that the reduced lag times actually fall on the relation followed by most other gluconeogenic substrates, as predicted (black curve in Fig. 3b).

To directly test the prediction of a linear relation between the inverse lag time and the abundance of lower gluconeogenic enzymes (equation (2)), we considered a shift from glucose to pyruvate, where a single gluconeogenic enzyme, phosphoenolpyruvate synthetase (PpsA), is required for the lower gluconeogenic reaction. We constructed a strain with linearly titratable PpsA expression that had a negligible effect on preshift growth. Titrating PpsA expression indeed affected the lag time of the glucose-to-pyruvate shift, and the model prediction was quantitatively validated by the observed proportionality between the preshift induction level of PpsA and the inverse lag time (Fig. 3c) over a fivefold range in lag times. As the full induction of PpsA in postshift alone was insufficient to overcome the lag phase, whereas preshift induction resulted in a large reduction in lag time, our results show the importance of expressing gluconeogenic enzymes in glycolytic conditions to shorten lag phase.

An important remaining question is why *E. coli* cannot avoid the depletion of gluconeogenic metabolite pools after shifting to gluconeogenesis. We hypothesized that allosteric regulation of the opposing glycolytic enzymes by metabolic intermediates does not achieve a complete inhibition of their activities during lag phase. To test whether residual activity of glycolytic enzymes may be a major cause of a long lag, we overexpressed glycolytic enzymes catalysing irreversible reactions in preshift conditions. Indeed, this severely impaired the switch from glycolysis to gluconeogenesis, more than doubling the lag time in most cases, as compared with preshift overexpression of a control enzyme (Extended Data Fig. 8). As glycolytic enzymes are abundant throughout the lag phase of the wild-type strain (Extended Data Fig. 6), the transition from glycolysis to gluconeogenesis is probably dominated by futile cycling, with both gluconeogenic and glycolytic enzymes active and working in opposite directions.

In this study, we have established a series of low-metabolite pools in gluconeogenesis as the cause of long lags during the transition from glycolysis. This is because, for fast glycolytic growth, the distribution of enzymes strongly favours glycolysis over the opposing gluconeogenesis (Extended Data Fig. 7c). At lower glycolytic fluxes, such as on poor glycolytic substrates, the change in the enzyme distribution (lower glycolytic enzyme and higher gluconeogenic enzyme abundances) favours glycolysis less, and the transition to gluconeogenic growth becomes faster. Thus, the two important fitness measures—growth rate and adaptability (inverse lag time)—are constrained as captured by equation (1). As this simple empirical relation holds broadly for many pairs of carbon sources tested (Fig. 1d and Extended Data Fig. 2), we propose that equation (1) be considered a phenomenological law of the growth–adaptation tradeoff, with the quantitative form arising from the structure of central carbon metabolism as suggested by the model in Extended Data Fig. 7.

The existence of this tradeoff suggests that it might be advantageous for cells to choose slower growth for the benefit of a shorter lag, in anticipation of switching to gluconeogenesis when the primary glycolytic substrates run out. It provides a unique perspective on the notorious problem of why bacteria grow on different substrates at broadly disparate rates. Hence, the quality of a substrate, as measured by growth rate, is a reflection of the ecological likelihood that conditions will change in fluctuating natural environments or across the bacterial infectious cycle, rather than on the basis of fundamental biochemical properties of the substrate, such as its energy content. As an example, wild-type *E. coli* grew substantially more slowly on fructose and mannose than on glucose, despite their similar chemical properties. Knocking out Cra—a transcriptional regulator that activates the expression of gluconeogenic enzymes while repressing those of glycolytic enzymes—increased growth on both fructose and mannose (Extended Data Fig. 9a), but was unable to support a shift to many gluconeogenic substrates. Thus, Cra may be designated to hold back the growth of wild-type cells on glycolytic substrates to enable a swift shift to gluconeogenesis when necessary. More notable is the growth on glycerol, often thought of as a poor nutrient compared with glucose owing to its reduced energy content. A single-residue mutation in the glycerol-uptake protein GlpK, which increases its uptake efficiency, accelerates growth on glycerol by more than 20% (refs. [16,17]). This faster-growing mutant has been extensively characterized[18], but a disadvantage of the mutation was only shown when combined with additional mutations[19], raising the possibility that *E. coli* may simply be maladapted to glycerol. Guided by our model, we find this mutant to exhibit a substantially longer lag compared with the slower-growing wild type (Fig. 3d), suggesting that slower growth of wild-type *E. coli* on glycerol might be selected to reduce the lag time upon abrupt transition to gluconeogenic substrates in the natural habitat.

This growth–adaptation tradeoff can be turned into a quantitative criterion for selecting the rate of cell growth (λ) by minimizing the total time for growth on a glycolytic substrate (roughly $1/\lambda$) together with its subsequent lag, $T_{\text{lag}}(\lambda)$ (equation (1)). Using parameters for the *E. coli* strain characterized here, and assuming that the environment provides glycolytic substrate at a concentration that would support bacterial growth by a factor of N (Extended Data Fig. 10a), we obtain an optimal glycolytic growth rate, λ^*, for which the time spent on growth and lag is balanced and minimized (Extended Data Fig. 10b).

Values for the optimal growth rate range from 0.5 h^{-1} to 1 h^{-1} for a broad range of nutrient abundances (Extended Data Fig. 10c), coinciding rather well with the range of growth rates observed for our strain on different glycolytic carbon sources[2]. The growth–adaptation tradeoff may thus be an important factor in the evolutionary selection of growth rate on specific substrates. As anaerobic bacteria typically do not grow on gluconeogenic carbon sources, they do not encounter these lag phases, and hence our model would predict selection of fast growth on many carbon sources. Indeed, we found that the gut anaerobe *Bacteroides thetaiotaomicron* grew at a similarly fast rate on several tested carbon sources (Extended Data Fig. 9e), indicating that the tradeoff did not play an important part in selecting their growth rates.

On the other hand, we do expect a similar tradeoff to exist in other respiro-fermentative microorganisms that are capable of growing on gluconeogenic carbon sources, because the biochemical structure of central metabolism is highly conserved. Indeed, we confirmed the existence of the tradeoff in the strictly aerobic bacterium *Bacillus subtilis* and in two wild-type strains of the single-celled eukaryote *Saccharomyces cerevisiae* (Extended Data Fig. 9b–d).

Recent studies have identified several conflicting objectives that affect microbial phenotypes[8–10,20–22]—for example, growth and motility[2,23,24], or growth and survival[25–27]. The establishment of quantitative relations for these and other pairs of conflicting traits could be expected to connect apparently disparate fitness measures into a unified framework. Identifying their occurrences and elucidating their origins will be crucial for gaining a better understanding of the diversity of microbial phenotypes across conditions and across species.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2505-4.

1. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
2. Basan, M. et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* **528**, 99–104 (2015).
3. Hui, S. et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol*. **11**, 784 (2015).
4. LaCroix, R. A. et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol*. **81**, 17–30 (2015).
5. Utrilla, J. et al. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. *Cell Syst*. **2**, 260–271 (2016).
6. O'Brien, E. J., Utrilla, J. & Palsson, B. O. Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. *PLOS Comput. Biol*. **12**, e1004998 (2016).
7. Towbin, B. D. et al. Optimality and sub-optimality in a bacterial growth law. *Nat. Commun*. **8**, 14123 (2017).
8. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604 (2012).
9. Shoval, O. et al. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157–1160 (2012).
10. Tendler, A., Mayo, A. & Alon, U. Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells. *BMC Syst. Biol*. **9**, 12 (2015).
11. Reimers, A.-M., Knoop, H., Bockmayr, A. & Steuer, R. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proc. Natl Acad. Sci. USA* **114**, E6457–E6465 (2017).
12. Kotte, O., Volkmer, B., Radzikowski, J. L. & Heinemann, M. Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. *Mol. Syst. Biol*. **10**, 736 (2014).
13. Baranyi, J. & Roberts, T. A. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol*. **23**, 277–294 (1994).
14. You, C. et al. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013).
15. Erickson, D. W. et al. A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* **551**, 119–123 (2017).
16. Zwaig, N. & Lin, E. C. C. Feedback inhibition of glycerol kinase, a catabolic enzyme in *Escherichia coli*. *Science* **153**, 755–757 (1966).
17. Pettigrew, D. W., Liu, W. Z., Holmes, C., Meadow, N. D. & Roseman, S. A single amino acid change in *Escherichia coli* glycerol kinase abolishes glucose control of glycerol utilization in vivo. *J. Bacteriol*. **178**, 2846–2852 (1996).
18. Lin, E. C. C. Glycerol dissimilation and its regulation in bacteria. *Annu. Rev. Microbiol*. **30**, 535–578 (1976).
19. Freedberg, W. B., Kistler, W. S. & Lin, E. C. Lethal synthesis of methylglyoxal by *Escherichia coli* during unregulated glycerol metabolism. *J. Bacteriol*. **108**, 137–144 (1971).
20. Goelzer, A. et al. Quantitative prediction of genome-wide resource allocation in bacteria. *Metab. Eng*. **32**, 232–243 (2015).
21. Maarleveld, T. R., Wortel, M. T., Olivier, B. G., Teusink, B. & Bruggeman, F. J. Interplay between constraints, objectives, and optimality for genome-scale stoichiometric models. *PLoS Comput. Biol*. **11**, e1004166 (2015).
22. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol*. **9**, 693 (2013).
23. Yi, X. & Dean, A. M. Phenotypic plasticity as an adaptation to a functional trade-off. *eLife* **5**, e19307 (2016).
24. Fraebel, D. T. et al. Environment determines evolutionary trajectory in a constrained phenotypic space. *eLife* **6**, e24669 (2017).
25. Vasi, F. K. & Lenski, R. E. Ecological strategies and fitness tradeoffs in *Escherichia coli* mutants adapted to prolonged starvation. *J. Genet*. **78**, 43–49 (1999).
26. Rozen, D. E., Philippe, N., Arjan de Visser, J., Lenski, R. E. & Schneider, D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecol. Lett*. **12**, 34–44 (2009).
27. Ying, B.-W. et al. Evolutionary consequence of a trade-off between growth and maintenance along with ribosomal damages. *PLoS ONE* **10**, e0135639 (2015).

Methods

Descriptions herein refer to a shift to acetate, but the same methods also apply to other shifts. No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Strain construction

All *E. coli* strains used here were derived from *E. coli* strain K-12 NCM3722 (ref. [28]). *B. thetaiotaomicron* was obtained from the American Type Culture Collection (ATCC catalogue number 29148).

P*tet*–*aceB* (NQ1350) and P*tet*–*ppsA* (NQ1357) strains. The DNA region containing the *km*^{r} gene, *rrnB*T and the promoter Ptet of the pKDT P*tet* plasmid[29] was amplified by polymerase chain reaction (PCR) with upstream and downstream primers ptet–aceB–insert (forward)/ptet–aceB–insert (reverse) and Ptet–ppsA–insert (forward)/Ptet–ppsA-insert (reverse), respectively, and then integrated into the chromosome of *E. coli* strain NQ309 to replace the chromosomal promoters of *aceB* and *ppsA* (each from −150 base pairs (bp) to −1 bp relative to the transcriptional start site)[30]. Each of the P*tet*–promoter substitutions was then transferred to strain NQ1358 (NCM3722 ycaD: P*tet*–*tetR* Δ*km*^{r})[29] backgrounds by phage P1 *vir*-mediated transduction, resulting in the strains NQ1350 and NQ1357.

***cra* deletion strain (NQ1077).** The Δ*cra* deletion allele in strain LJ2801 (*E. coli* Genetic Stock Center), in which a *km*^{r} gene is substituted for *cra*, was transferred to wild-type strain NCM3722, resulting in strain NQ1077.

PykF (NQ1543), PfkA (NQ1544) and ArgA (NQ1545) overexpression strains. Overexpression pNT3 plasmids (from the library in ref. [31]) expressing the genes *pykF*, *pfkA* or *argA* from plasmid P*tac* were purified and transformed into wild-type strain NCM3722, resulting in the strains NQ1543, NQ1544 and NQ1545, respectively.

***glpK22* mutant strain (NQ898).** To create a strain that grows faster on glycerol, we replaced the *glpK* gene in strain NCM3722 with the *glpK22* variant[17] through two P1 transduction steps. First, the *pfkA*::km marker was transferred into strain NCM3722 using phage P1 *vir*, prepared from the Keio collection[32]. The resulting strain (NQ632) from the transduction cannot use mannitol as its sole carbon source. Second, phage P1 *vir* prepared from strain CGSC5511 (Lin-43)[16] containing *glpK22* was transfected into NQ632. Selecting a colony that grew on mannitol minimum medium yielded a strain, NQ898, containing the *glpK22* mutation in an NCM3722 background.

YCE44 strain. The recipient strain NCM3722 was used for P1 transduction[33] with P1 lysate prepared from the Keio collection[32] to create the *fliC*::Kan mutant. This mutant was then transformed with the Pcp20 plasmid[30] to flip out the kanamycin marker. The resulting strain was then used as a recipient strain for P1 transduction with BO37 (ref. [34]) P1 lysate to create the final target strain, YCE44 (NCM3722, *fliC*::FRT–FRT, *glmS*::PRNAI–*mCherry1*–11–*mKate*–T1 terminator–FRT Kan *FRT*::*pstS*). The donor strain BO37 was provided by the Paulsson Laboratory[34].

Strains used herein. Except for wild-type strain BW25113, used as a control, all strains herein were derived from *E. coli* K-12 strain NCM3722 (refs. [28,35,36]), provided by the S. Kustu laboratory. See Supplementary Table S1 for a summary of strains.

Growth of bacterial cultures

Growth media. Unless otherwise indicated, we used N$^+$C$^+$ minimal medium[37], which contains K_2SO_4 (1 g), $K_2HPO_4.3H_2O$ (17.7 g), KH_2PO_4 (4.7 g), $MgSO_4.7H_2O$ (0.1 g) and NaCl (2.5 g) in one litre, and is supplemented with 20 mM NH_4Cl and specified carbon sources. Carbon-source concentrations were based on the number of carbon atoms in the molecule: 20 mM for C_6 carbons, 30 mM for C_4 carbons and 40 mM for C_3 carbons.

The base minimal medium used for the anaerobic growth of *E. coli* strain NCM3722 consisted of KH_2PO_4 (2 g), K_2HPO_4 (14.8 g), NaCl (0.58 g), NH_4Cl (0.54 g) and Na_2SO_4 (0.07 g), and 1,000× mineral solution (1 ml) per litre. One litre of the ×1,000 mineral solution contained $MgCl_2$ (60 g), $CaCl_2$ (5.5 g), $FeSO_4.7H_2O$ (5.5 g), $MnCl_2$ (19.7 mg), $CoCl_2$ (23.8 mg), Ni_2SO_4 (26.2 mg), $CuCl_2$ (15.9 mg), $(NH_4)_2MoO_4$ (23.5 mg), SeO_2 (11 mg), $ZnSO_4$ (28.7 mg) and H_3BO_4 (6.2 mg) dissolved in 100 mM HCl. For consistency of comparisons, the same medium was used for the aerobic growth of strain NCM3722 (Extended Data Fig. 9). Carbon sources were added as indicated.

The medium used for the anaerobic growth of *B. thetaiotaomicron* was the same as that used for the anaerobic growth of *E. coli*, but also included 2 mg cyanocobalamin, 2 mg haemin and 0.6 g cysteine per litre. For anoxic media, Hungate tubes (16 mm × 125 mm) filled with 7 ml medium were shaken at 270 rpm under a 7% CO_2, 93% N_2 atmosphere pressurized to 1.5 atm for 75 min. Cultures were transferred anoxically into Hungate tubes with disposable syringes.

Growth measurements. Batch culture aerobic growth was performed in a 37 °C water-bath shaker or air incubator shaking at 250 rpm. The culture volume was at most 10 ml in 25 mm × 150 mm test tubes. For seed culture, one colony from fresh LB agar plates was inoculated into liquid LB and cultured at 37 °C with shaking. Cells were then diluted into minimal medium and cultured in a 37 °C water-bath shaker overnight (preculture). The overnight preculture was allowed to grow for at least three doublings. Cells from the overnight preculture were then diluted to OD_{600} = 0.005–0.025 in identical prewarmed minimal medium, and cultured in 37 °C water bath shaker (experimental culture). At every half-doubling, we collected 200 µl of cell culture in a Sterna submicrometre cuvette for OD_{600} measurement using a thermal spectrophotometer, after allowing at least four generations of growth. The time taken for each sample collection was less than 30 s and had no measureable effect on cell growth.

Anaerobic growth was performed similarly with a few exceptions. All growth for *B. thetaiotaomicron* was carried out in Hungate tubes. For seed culture, a single colony from Wilkens–Chalgren agar plates was inoculated into anoxic Hungate tubes filled with 7 ml Wilkens–Chalgren broth and incubated at 37 °C with shaking. Cells were then diluted roughly 300-fold into preculture medium to grow overnight. The next day, cells were diluted to OD_{600} = 0.01–0.025 for experimental cultures in the same medium as the preculture. OD_{600} measurements of cells in Hungate tubes were made with a Thermo Genesys 20 modified to hold Hungate tubes in place of a cuvette. To maintain the temperature of the culture tube, we removed tubes from the water-bath shakers to measure OD_{600} and returned them within 30 s. The OD_{600} measured through the Hungate tubes was equivalent to the OD_{600} measured through a cuvette in the range of at least 0.04–0.5.

Anaerobic growth of *E. coli* strain NCM3722 was measured similarly to that of *B. thetaiotaomicron*, except that seed cultures were performed aerobically in LB broth before being diluted roughly 300-fold into anoxic Hungate tubes for overnight preculture with the same media as the experimental culture. Cells were again diluted into fresh Hungate tubes with OD_{600} = 0.01–0.025 for experimental culture, and growth was measured with the modified Thermo Genesys 20.

pH changes. Because anaerobic growth of *E. coli* and *B. thetaiotaomicron* involves copious acid production, the pH of cultures was monitored. Typical pH changes for the anaerobic growth of NCM3722 were from 7.2 (fresh anoxic medium) to 6.7 (at an OD_{600} of roughly 0.4). For

B. thetaiotaomicron, the pH changes were from 7.2 (fresh anoxic medium) to 6.9 (OD_{600} approximately 0.4). The pH for the aerobic growth of NCM3722 stayed at around 7.4–7.3 for fresh medium and cultures at an OD_{600} of roughly 0.4.

Medium shift and determination of lag times

***E. coli* growth.** Exponentially growing cultures in the preshift condition were obtained as above, in tubes or in flasks. For metabolomics and proteomics experiments, cultures were grown to a higher OD_{600} of approximately 0.5 before the shift was performed. Cells were then carefully transferred to a filter (previously washed with Milli-Q water) to remove preshift medium and washed twice with warmed postshift medium (at least twofold the volume of culture transferred to the filter). The filter was then moved to a sterile 50 ml tube with warmed postshift medium, and cells were gently resuspended from the filter by pipetting. Cells were then diluted in warmed postshift medium to an OD_{600} of roughly 0.05 for the purpose of lag-time measurements, and of roughly 0.5 for the purpose of metabolomics and proteomics measurements, and incubated. The entire shift was typically completed in under 5 min. Lag times were determined as follows. After cells reached steady-state growth in the postshift condition, about three to four OD_{600} data points were fitted with an exponential function. The intersection of the fitted exponential and initial postshift OD_{600} was used to determine the lag time.

To screen combinations of carbon sources using a plate reader, we modified the protocol slightly. After being transferred to the filter, cells were washed twice and resuspended using warmed medium without a carbon source. Cells were then diluted into prewarmed Thermo Fisher Scientific Nuclon 96-well flat-bottom transparent plates filled with different postshift media. These plates were covered with lids and incubated; culture density was monitored using a Tecan Infinite M200 plate reader at 37 °C, shaking at 880 rpm, to measure lag times. Lag times were determined by fitting the growth curve over the range in which the maximal exponential growth rate was reached, using the function $OD(t) = OD_{init} \exp[\lambda(t - T_{lag})]$, which is an exponential growth curve with growth rate λ shifted by lag time T_{lag}. OD_{init} is the OD_{600} measured just after the shift, and the fit parameters were λ and T_{lag}. The fit was performed using the 'fit' command of Gnuplot, which is an implementation of the nonlinear least-squares (NLLS) Marquardt–Levenberg algorithm.

***B. subtilis* growth.** A single colony of *B. subtilis* strain 3610 was inoculated in 3 ml LB broth in the morning as a seed culture at 37 °C. In the evening, the seed culture was diluted into minimal medium containing various carbon sources (20 mM glucose, 20 mM mannose, 20 mM maltose and 40 mM glycerol) to ensure exponential growth the next day. The seed culture was then diluted to an OD_{600} of 0.025. When the culture reached an OD_{600} of 0.2–0.3, cells were centrifuged, washed with prewarmed postshift medium, and shifted to minimal medium containing 60 mM acetate. OD values were recorded using a BioTek Synergy H1 microplate reader.

Yeast growth. Overnight seed cultures of *S. cerevisiae* strains YPS128 and YPS163 were grown in chemically defined synthetic complete media[38–40], containing 2% (*w*/*v*) glucose. The next day, the seed culture was diluted to an OD_{600} of 0.025 in synthetic complete medium containing various single carbon sources, namely 2% (*w*/*v*) of glucose, galactose, maltose or raffinose, and incubated at 30 °C. When cultures reached the exponential phase (an OD_{600} of 0.2–0.4), cells were washed twice with prewarmed postshift medium and shifted to the postshift medium containing 2% (*w*/*v*) acetate. Growth was followed and OD_{600} values were recorded using a BioTek Synergy H1 microplate reader. The chemically defined synthetic complete media used for this yeast carbon switch experiment left out inositol completely to ensure that cells were growing on a single carbon source.

Mother-machine methods. We used a microfluidic platform based on the 'mother machine' design[43] to track individual cells during lag phase. We monitored the morphology of individual cells as they experienced a switch in medium under controlled conditions, and used the morphological measurements to obtain both growth rates and lag times of individual cells (single-cell lag-time analysis).

The mother-machine microfluidic device, in which cells grow and divide in narrow trenches and are fed through diffusion by an orthogonal feeding channel, has been used for long-term tracking of cells[41,42] under tightly controlled local conditions. The Paulsson Laboratory has recently applied[43] this microfluidic platform to the tracking of cell lineages in different phases of the growth curve using a new setup, in which a batch culture is directly connected to the microfluidic chip. We used this platform here to obtain lag-time information at the single-cell level (Extended Data Fig. 3a). Cells from the YCE44 strain (constitutively expressing mCherry1-11-mKate) were loaded onto a mother-machine chip and were allowed to recover for several hours in N^+C^+ glucose minimal medium before starting imaging. A flask with glucose medium inoculated with the YCE strain[44] was then connected to the microfluidic device, so that the cells in the chip shared the same environment as the cells in the flask. The platform enabled us to monitor the optical density of the batch culture at high frequency (30 s), and to grow the culture under usual laboratory conditions (37 °C on an orbital shaker, 220 rpm). This allowed us to monitor the behaviour of the batch culture and individual cells synchronously. To perform the shift to acetate, cells in the flask were washed twice with postshift acetate minimal medium and resuspended in postshift acetate minimal medium as described in 'Growth measurements'. After the shift, cells kept growing for some time at the same growth rate both in the flask and in the microfluidic chip, probably because some glucose medium was still present in the system. After about 60 min, the glucose ran out and the cells underwent a diauxic shift. We kept monitoring cells in the mother machine over the course of the lag phase, as they responded to changes in the batch culture. The experimental protocol is illustrated in Extended Data Fig. 3b.

Cell conditions in the microfluidic chip were not identical to those in the flask: for instance, cells under observation were diffusely fed in the growth trenches. We minimized this effect by using shorter growth trenches (20 µm in length). Also, in order to reduce mixing of glucose and acetate media at the time of switch, we introduced a waste line before the microfluidic chip, which allowed us to divert the flow at the time of switching, to better control the switch dynamics for the cells in the mother machine.

Imaging parameters. Images were acquired using a Nikon titanium inverted microscope equipped with a temperature-controlled incubator (Okolab), an Andor Zyla 4.2 camera, a ×40 phase 2 Plan Apo objective (numerical aperture (NA) 0.95, Nikon), an automated motorized stage (Nikon) and a Lumencor SpectraX light engine (https://lumencor.com/products/spectra-x-light-engine/). All images were acquired with ×1.5 post-magnification, and the camera–objective combination gave a resolution of 0.11 µm $pixel^{-1}$. Focal drift was controlled by the Nikon Perfect Focus System. The timelapse imaging and automatic stage movements were controlled by Nikon NIS Elements software. We imaged cells in phase contrast and red fluorescent protein (RFP) channel. Images were taken every 15 min with an exposure of 200 ms in order to reduce photobleaching and phototoxicity.

Image analysis pipeline

Segmentation (FIJI). After trying a few segmentation approaches using either FIJI or Python, we opted for using a custom FIJI macro in combination with manual selection of trenches. Individual lineages were selected before segmentation, and trenches with double-loading (where cells were loaded side-by-side in a growth trench and grew under stressful conditions and poor feeding) or which were out of focus were discarded. Out of a total of 1,494 starting trenches, 363 presented double loading and 44 became unloaded; 7 mother cells did not wake up after the switch to acetate, 2 cells lysed after the switch, and 114 cells were discarded for various reasons (for example, they were out of focus or not growing at all before the shift the acetate). The remaining 964 cells were segmented using the fluorescence channel (RFP) with a custom FIJI algorithm based on thresholding, morphological transformations and an adjustable watershed, designed to work for cells with changing sizes (cells substantially change their morphology between glucose and acetate media and along the growth curve). We then proceeded to inspect each mask produced, to discard trenches with too many visible segmentation errors that might affect the single-cell lag-time analysis. Of the 964 trenches segmented, we selected 685 with near-perfect segmentation.

Analysis (Matlab). We focused solely on those cells at the top of the growth trenches ('mother cells'), as we could follow them for the entire experiment and extract single-cell traces for the full duration. The temporal information of the cell data (such as length and area) was then compiled into single-cell length traces. We identified cell divisions by using a findpeaks package, looking at sudden decreases in cell length but still filtering out fluctuations from segmentation mistakes.

From a total of 685 mother cells with near-perfect segmentation, we removed three cells that had missing measurements along the time trace or became unloaded from the microfluidic trench. We checked also for cells with no divisions during experiments or after the switch, for filamenting cells (longer than 8 µm) and for cells not recovering after the switch. One cell exhibited filamentation and we proceeded with analysis of the remaining 681 cells.

We estimated that the medium should flow through the microfluidic chip at around frame 47 (11.75 h from the start of imaging). In order to confirm this determination of the time of the switch to acetate medium in the mother machine microfluidic chip, we used the single-cell instantaneous growth rate. Cells started to slow their growth at frame 47 (11.75 h), and globally reached a minimum at frame 50 (12.50 h). In the rest of the analysis, we used frame 47 as the switching time to acetate medium and frame 50 as starting time for the lag-time computation.

To compute the lag time for each individual cell, we needed to compute the growth rate at the single-cell level. We used the instantaneous growth rates of individual cells determined from changes in cell length between adjacent time points for each birth-to-division event (Extended Data Fig. 3c). The lag time for each individual cell could then be computed using the following formula:

$$T_i^{\mathrm{lag}}(t) = t - \frac{1}{\lambda_i^{\mathrm{ACE}}}\int_0^t \lambda_i(t)dt$$

in which $\lambda_i(t)$ is the instantaneous growth rate of cell i at time t, and λ_i^{ACE} is the maximum growth rate that cell i attains in acetate medium. We used the time of minimum growth rate for the population (frame 50) as the starting point for computation of the lag time (time 0 in previous formula). The lag time from the equation above is a monotonically increasing function of time, and it reaches a plateau when the growth rate approaches λ_i^{ACE}. This plateau corresponds to the single-cell lag time; the resulting distribution is shown in Extended Data Fig. 3d (one of the cells was removed from the analysis as it did not wake up in acetate medium; the analysis was performed on a total of 680 cells).

Using the mother machine, we followed the initial population of cells loaded into the device. However, variability in growth of individual lineages must be considered when comparing results from mother-machine data at the population level with results from the batch culture, as cells in the mother machine are not subjected to the dilution effect that occurs in batch. Assuming that the progeny of each cell in the mother machine maintain the same growth characteristics as that progenitor cell, and assuming the same initial cell size, we can calculate the expected batch dynamics from the single-cell data in the mother machine. If we denote with $\lambda_i(t)$ the growth rate of cell i in the mother machine at time t, and if $\lambda_{\mathrm{b}}(t)$ is the instantaneous growth rate of the batch population, then the normalized batch OD_{600} is given by:

$$\mathrm{OD}_{\mathrm{b}}(t)/\mathrm{OD}_{\mathrm{b}}(0) = \exp\left(\int_0^t \lambda_{\mathrm{b}}(s)\,ds\right) = \frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(\int_0^t \lambda_i(s)\,ds\right)$$

in which N_0 is the number of cells that we observe in the mother machine, and time 0 is the time when population attained a minimum in growth rate (frame 50). This equation can be used to calculate the batch growth rate, $\lambda_{\mathrm{b}}(t)$, from single-cell data and to derive the expected lag time for the batch culture, $T_{\mathrm{b}}^{\mathrm{lag}}(t)$:

$$T_{\mathrm{b}}^{\mathrm{lag}}(t) = t - \frac{1}{\lambda_{\mathrm{b}}^{\mathrm{ACE}}}\log\left[\frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(\int_0^t \lambda_i(s)\,ds\right)\right]$$

in which $\lambda_{\mathrm{b}}^{\mathrm{ACE}}$ is the maximum of the expected batch growth rate, $\lambda_{\mathrm{b}}(t)$, in acetate medium, and the integral is performed to the time point where $\lambda_{\mathrm{b}}(t) = \lambda_{\mathrm{b}}^{\mathrm{ACE}}$. When the growth rate reaches its steady-state, $T_{\mathrm{b}}^{\mathrm{lag}}(t)$ is invariant for different integration times, t.

Because the experimental setup includes high-frequency OD_{600} measurements (30 s interval) of the connected batch culture flask (Extended Data Fig. 3e), we could use these data to compute the batch lag time and have a direct comparison between the batch culture and the single-cell data. Similarly to the previous formula, the lag time for the batch culture could be computed using the formula:

$$T^{\mathrm{lag}}(t) = t - \frac{1}{\lambda_{\mathrm{ACE}}}\log\left(\frac{\mathrm{OD}(t)}{\mathrm{OD}(0)}\right)$$

in which λ_{ACE} corresponds to the maximum growth rate in acetate medium, and $t = 0$ is the time at which the bulk culture halts growth after the switch to acetate. The lag time of the batch corresponds to the value of $T^{\mathrm{lag}}(t)$ at which the growth rate in the flask approaches λ_{ACE}, which corresponds to a plateau for the function $T^{\mathrm{lag}}(t)$.

Batch microscopy

Experimental protocol. NCM3722 wild-type cells were grown in $\mathrm{N^+C^+}$ glucose medium as above. When the batch culture reached an OD_{600} of 0.2, cells were collected by filtering and washed twice in $\mathrm{N^+C^+}$ acetate medium (as in all other medium-shift experiments described above). After the washing step, cells were resuspended in $\mathrm{N^+C^+}$ acetate medium to reach a final OD_{600} of 0.05. This culture was split into two identical six-well glass-bottom plate (Cellvis, number 1.5), with 5 ml of culture in each well. One of the six-well plates was centrifuged at 4,800*g* for 3 min and bacterial cells were imaged on a Nikon Ti2 microscope (×40 air phase contrast objective). The plate was kept for imaging in a 37 °C temperature-controlled microscope chamber. Phase-contrast images were taken from multiple fields of views with a frame rate of 300 seconds. The other six-well plate was taken to a shaker air incubator (kept at 37 °C and 220 rpm). This plate was considered to be the batch culture control. We measured OD_{600} and calculated the batch lag time (295 min) from the recorded optical density measurements (Extended Data Fig. 4).

Analysis of microscopy data. After recording the microscopy data, we carried out image analysis using a custom analysis pipeline in Python. In brief, each time series was first corrected for *XY* drift using rigid body transformations[44]. After drift correction, single-cell time traces were segmented using Otsu thresholding. Cell tracing stopped when the cell divided, or the field of view became obstructed by adjacent dividing cells, or the cell became dislodged from the glass surface and we lost track of it. Cells that were followed for 43 or more frames were considered for analysis. This threshold was chosen on the basis of a systematic analysis of different values for this threshold. We wanted to establish an upper bound on the number of nongrowing cells after the shift to acetate. We did not expect nongrowing cells to be overrepresented in transiently present cells that briefly settled on the glass bottom and were then washed away. These transiently present cells become more important for low values of the threshold. On the other hand, for high values of the threshold we were artificially enriching for non-growing cells. The intermediate value that we chose established the most stringent upper bound for the fraction of nongrowing cells in the population.

We segmented 1,761 cells, after which we set an arbitrary threshold of a 10% increase in single-cell area to identify cells that showed significant growth. In Extended Data Fig. 4, cell traces that crossed this 10% threshold are marked in blue, and cells that did not are marked in red. Out of 1,761 segmented single-cell traces, 1,500 (roughly 85.17%) crossed the chosen 10% threshold, and only 261 (around 14.83%) showed less than a 10% increase in area over the experiment (Extended Data Fig. 4). Therefore, using our method we have detected 14.8% non-growing cells. This number sets an upper bound to the fraction of the non-growing cell population. It is likely that many of these cells would have showed substantial growth at later time points, which we were unable to measure owing to experimental limitations. This suggests that the actual population of cells that do not resume growth is in reality much smaller than the roughly 14.8% that we have measured.

Metabolite mass spectrometry

Sample collection and quenching. For metabolite measurements and ^{13}C-labelling experiments, we transferred an amount proportional to 1 ml*OD$_{600}$ of the culture broth onto a Durapore filter with a pore size of 0.45 μm (Millipore) and vacuum-filtered the sample. For metabolite measurements, the filter with cells was immediately transferred after filtration into 4 ml of 20 °C acetonitrile/methanol/water (2/2/1) to quench metabolism and 200 μl of a uniformly ^{13}C-labelled *E. coli* metabolite extract were added as an internal standard[45]. ^{13}C-labelling experiments were performed immediately after vacuum-filtration on the filter, as described[46]. Specifically, cells on the filter were first washed with fresh, preheated (37 °C) acetate M9 medium for 10 s, and ^{13}C-labelling was initiated by changing the washing solution to preheated (37 °C) M9 medium containing uniformly ^{13}C-labelled acetate. After each labelling step, the filter was transferred into 4 ml of 20 °C acetonitrile/methanol/water (2/2/1) for quenching. To extract metabolic intermediates, the filter was kept in this solution at −20 °C for 1 h. Then the cell debris was removed from the extracts by centrifugation (4 °C, 10,000 rpm, 10 min); the supernatants were transferred into new tubes and dried to complete dryness.

Sample preparation. For liquid chromatography/mass spectrometry (LC/MS) analysis, dried extracts were resuspended in 100 μl deionized water, of which 10 μl were injected into a Waters Acquity ultraperformance liquid chromatography (UPLC) system (Waters) with a Waters Acquity T3 column coupled to a Thermo TSQ Quantum Ultra triple quadrupole instrument (Thermo Fisher Scientific) with negative-mode electrospray ionization. Compound separation was achieved using a gradient of two mobile phases: A, 10 mM tributylamine, 15 mM acetic acid and 5% (*v*/*v*) methanol; and B, 2-propanol[47]. Mass isotopomer distributions of carbon backbones was acquired as described[48]. We carried out peak integration using in-house software (B. Begemann and N. Zamboni, personal communication).

Kinetic flux estimation. Flux estimation closely followed ref. [49], based on the kinetics of incorporation of a ^{13}C-acetate isotope. At numerous time points, after cells were rapidly switched from unlabelled to isotope-labelled acetate, LC/MS analysis was performed. Resulting plots of unlabelled compound versus time were fitted by an exponential decay, and the flux was calculated as the decay rate multiplied by the intracellular metabolite concentration.

Proteomic mass spectrometry

Metabolic labelling with ^{15}N (ref. [50]) provides relative quantitation of unlabelled proteins with respect to labelled proteins across growth conditions of interest. Each experimental sample in a series is mixed in an equal amount with a known labelled standard sample as reference, and the relative change of protein expression in the experimental sample is obtained for each protein.

Sample collection. For each culture, 1.8 ml of cell culture at OD$_{600}$ = 0.4–0.5 was collected by centrifugation. The cell pellet was resuspended in 0.2 ml of water and fast frozen on dry ice.

Sample preparation. A balanced mixture of the two ^{15}N-labelled cell samples (from glycolytic and gluconeogenic growth conditions, with cells grown on glucose and acetate respectively) was prepared as a universal reference. We added 100 μg of the labelled reference proteome to 100 μg of each experimental sample. This balanced preparation (equal amounts of total protein) enabled the measurement of proteome mass fraction for each protein. The mixed reference ensured that the distribution of proteins in the reference was not strongly biased by a particular growth condition.

Proteins were precipitated by adding 100% (*w*/*v*) trichloroacetic acid (TCA) to a final concentration of 25%. Samples were left to stand on ice for a minimum of 1 h. The protein precipitates were spun down by centrifugation at 13,200*g* for 15 min at 4 °C. The supernatant was removed and the pellets were washed with cold acetone and dried in a Speed-Vac concentrator.

The pellets were dissolved in 80 μl 100 mM NH_4HCO_3 with 5% acetonitrile (ACN). We added 8 μl of 50 mM dithiothreitol (DTT) to reduce the disulfide bonds before the samples were incubated at 65 °C for 10 min. Cysteine residues were modified by adding 8 μl of 100 mM iodoacetamide (IAA) followed by incubation at 30 °C for 30 min in the dark. Proteolytic digestion was carried out by adding 8 μl of 0.1 μg μl^{-1} trypsin (Sigma-Aldrich) with incubation overnight at 37 °C. The peptide solutions were cleaned by using PepClean C-18 spin columns (Pierce, Rockford, IL). After drying in a Speed-Vac concentrator, the peptides were dissolved into 10 μl sample buffer (5% ACN and 0.1% formic acid).

Mass spectrometry. The peptide samples were analysed on an AB SCIEX TripleTOF 5600 system (AB Sciex) coupled to an Eksigent NanoLC Ultra system (Eksigent). The samples (2 μl) were injected using an autosampler. The samples were first loaded onto a Nano cHiPLC Trap column (200 μm × 0.5 mm, ChromXP C18-CL, 3 μm, 120 Å; Eksigent) at a flow rate of 2 μl min^{-1} for 10 min. The peptides were then separated on a Nano cHiPLC column (75 μm × 15 cm, ChromXP C18-CL, 3 μm, 120 Å; Eksigent) using a 120-min linear gradient of 5–35% ACN in 0.1% formic acid at a flow rate of 300 nl min^{-1}. Settings were: MS1, mass range *m*/*z* 400–1,250 and accumulation time 0.5 s; MS2, mass range *m*/*z* 100–1,800, accumulation time 0.05 s, high sensitivity mode, charge state 2–5, selecting anything over 100 counts per second, maximum number of candidates per cycle 50, and excluding former targets for 12 s after each occurrence.

Protein identification. The raw mass spectrometry data files generated by the AB SCIEX TripleTOF 5600 system were converted to centroided mzml files, which were searched using the X!Tandem search engine (https://thegpm.org) against the *E. coli* proteome database (Uniprot; https://www.uniprot.org) to identify proteins. The following parameters were used in the X!Tandem searches: parent mass error 50 ppm, fragment mass error 100 ppm. Ions with charge 1, 5, 6 or 7 were ignored, as were peptides of fewer than six residues. Spectral libraries for each each condition were built and refined using Spectrast (ISB), only keeping those peptides that were identified in three or more individual samples, and collapsing individual spectra into a consensus spectra for each peptide.

Relative protein quantification. The raw mass spectrometry data files were converted to the mzml format using conversion tools provided by AB Sciex, and the consensus libraries from Spectrast were used to quantify each of the (non-centroided) mzml files using our in-house quantification software[8] (Massacre). In brief, the intensity for each peptide is integrated over a patch in *RT*, *m*/*z* space that encloses the envelope for the light and heavy peaks. After collapsing data in the *RT* dimension, the light and heavy peaks are fit to a multinomial distribution (a function of the chemical formula of each peptide) using a least squares Fourier transform convolution routine[9], which yields the relative intensity of the light and heavy species. The ratio of the unlabelled to labelled peak intensity is obtained for each peptide in each sample. A confidence measure for each fit is calculated from a support vector machine (SVM) trained on a large set of user scoring events.

The relative protein level for each protein in each sample is obtained as a ratio by taking the weighted median (using the SVM score) of the ratios of all its corresponding peptides.

Uncertainty of individual measurements

Biological replicates show the following typical uncertainties in measured quantities: growth rate, roughly 5%; lag times, roughly 15% for long lag times (longer than 1 h). Short lag times (less than 1 h) show higher relative variabilities.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Lag times are provided in Supplementary Tables 2, 3. All other data are found in downloadable Excel files for each figure. Data for Fig. 3a were taken from ref. [3] and are deposited with the paper on the *Molecular Systems Biology* website. Source data are provided with this paper.

28. Brown, S. D. & Jun, S. Complete genome sequence of *Escherichia coli* NCM3722. *Genome Announc*. **3**, e00879-15 (2015).
29. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366–1375 (2009).
30. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
31. Saka, K. et al. A complete set of *Escherichia coli* open reading frames in mobile plasmids facilitating genetic studies. *DNA Res*. **12**, 63–68 (2005).
32. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol*. **2**, 2006.0008 (2006).
33. Thomason, L. C., Costantino, N. & Court, D. L. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol*. 1.17.1–1.17.8 (2007).
34. Okumus, B. et al. Mechanical slowing-down of cytoplasmic diffusion allows *in vivo* counting of proteins in individual cells. *Nat. Commun*. **7**, 11641 (2016).
35. Soupene, E. et al. Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. *J. Bacteriol*. **185**, 5611–5626 (2003).
36. Lyons, E., Freeling, M., Kustu, S. & Inwood, W. Using genomic sequencing for classical genetics in E. coli K12. *PLoS ONE* **6**, e16717 (2011).
37. Csonka, L. N., Ikeda, T. P., Fletcher, S. A. & Kustu, S. The accumulation of glutamate is necessary for optimal growth of *Salmonella typhimurium* in media of high osmolality but not induction of the proU operon. *J. Bacteriol*. **176**, 6324–6333 (1994).
38. Hirsch, J. P. & Henry, S. A. Expression of the *Saccharomyces cerevisiae* inositol-1-phosphate synthase (INO1) gene is regulated by factors that affect phospholipid synthesis. *Mol. Cell. Biol*. **6**, 3320–3328 (1986).
39. Dowd, S. R., Bier, M. E. & Patton-Vogt, J. L. Turnover of phosphatidylcholine in *Saccharomyces cerevisiae*. The role of the CDP-choline pathway. *J. Biol. Chem*. **276**, 3756–3763 (2001).
40. Jesch, S. A., Zhao, X., Wells, M. T. & Henry, S. A. Genome-wide analysis reveals inositol, not choline, as the major effector of Ino2p-Ino4p and unfolded protein response target gene expression in yeast. *J. Biol. Chem*. **280**, 9106–9118 (2005).
41. Wang, P. et al. Robust growth of *Escherichia coli*. *Curr. Biol*. **20**, 1099–1103 (2010).
42. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503**, 481–486 (2013).
43. Bakshi, S. et al. Dynamic regulation of growth and physiology of microbes under complex changing conditions. Preprint at https://www.biorxiv.org/content/10.1101/2020.03.27.006403v2 (2020).
44. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process*. **7**, 27–41 (1998).
45. Wu, L. et al. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards. *Anal. Biochem*. **336**, 164–171 (2005).
46. Hörl, M., Schnidder, J., Sauer, U. & Zamboni, N. Non-stationary ^{13}C-metabolic flux ratio analysis. *Biotechnol. Bioeng*. **110**, 3164–3176 (2013).
47. Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem*. **82**, 4403–4412 (2010).
48. Rühl, M. et al. Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of ^{13}C metabolic flux analysis. *Biotechnol. Bioeng*. **109**, 763–771 (2012).
49. Yuan, J., Bennett, B. D. & Rabinowitz, J. D. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nat. Protocols* **3**, 1328–1340 (2008).
50. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**, 6591–6596 (1999).
51. Fenton, A. W. & Reinhart, G. D. Disentangling the web of allosteric communication in a homotetramer: heterotropic inhibition in phosphofructokinase from *Escherichia coli*. *Biochemistry* **48**, 12323–12328 (2009).

Acknowledgements We thank A. Murray for helpful comments and suggestions, and V. Patsalo for technical support and development of the proteomics method. M.B. acknowledges a SystemsX.ch fellowship. T. Honda acknowledges a Japan Student Services Organization (JASSO) long-term graduate fellowship award. Work in the Hwa laboratory is supported by the National Institutes of Health (NIH) through grant R01GM109069 and by the Simons Foundation through grant 330378. J.R.W. acknowledges NIH support through grant R01GM118850.

Author contributions M.B., T. Hwa and U.S. designed the study. Experiments were performed by M.B., T. Honda, M.H., Y.-F.C., E.L., A.M., H.O., B.R.T., J.M.S. and C.S., and all authors contributed to the analysis of experimental data. Specifically, lag times for *E. coli* were measured by M.B. and T. Honda. Metabolite measurements and analysis were performed by M.B. and M.H. Proteomics measurements were performed by T. Honda and H.O. Proteomics data analysis was performed by T. Honda, D.C., J.M.S and J.R.W. Genetic constructs were made by M.B., Y.-F.C and H.O. Lag times for *S. cerevisiae* and *B. subtilis* were measured by Y.-F.C. Growth rates for *B. thetaiotaomicron* were measured by B.R.T. Experiments for single-cell lag phases from microfluidics and plates were performed and analysed by E.L., A.M. and C.S. M.B., D.C., T. Hwa and U.S. developed the model. M.B., J.P., T. Hwa and U.S. wrote the paper and the Supplementary Information.

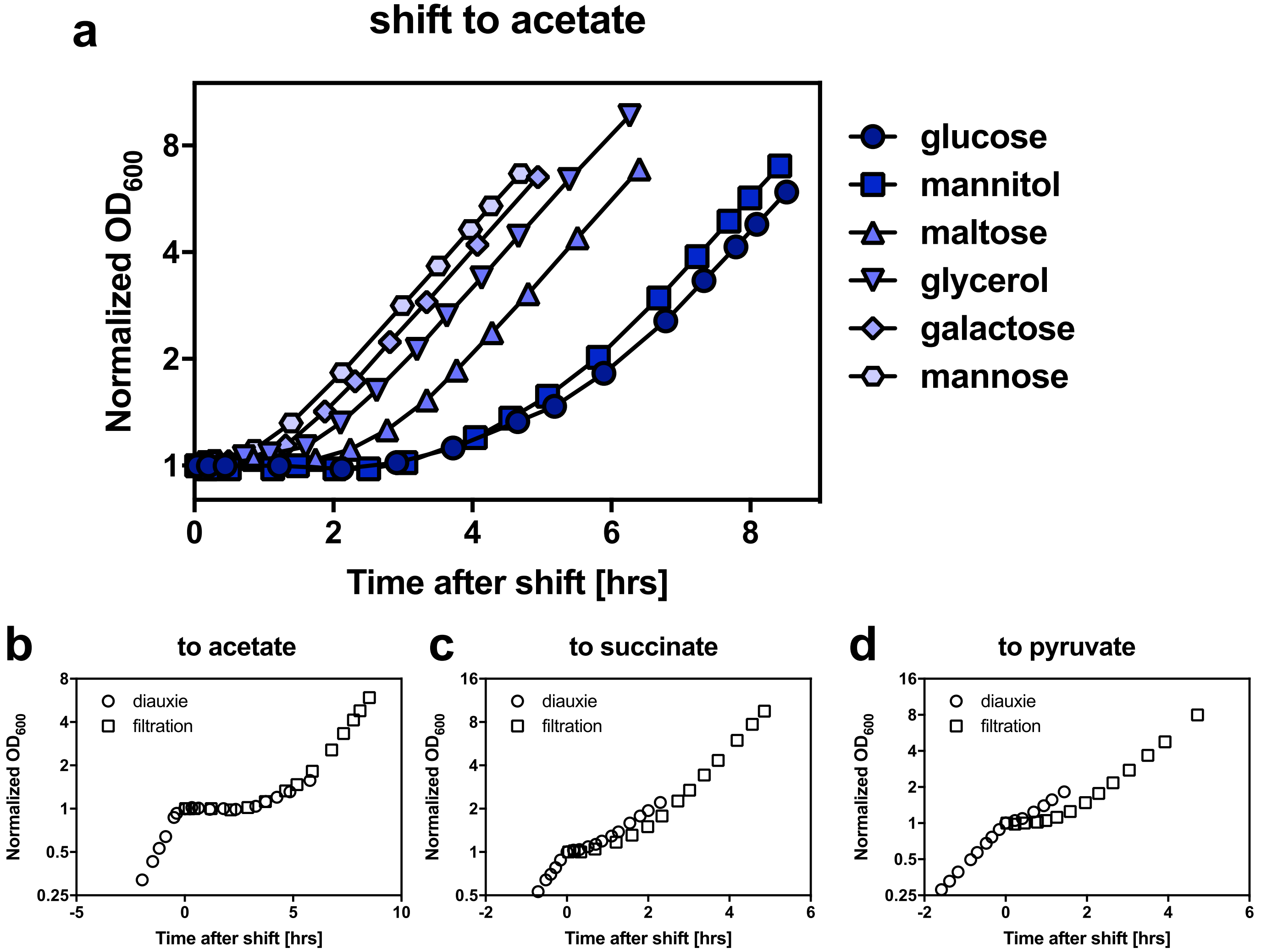
Competing interests The authors declare no competing interests.

Additional information

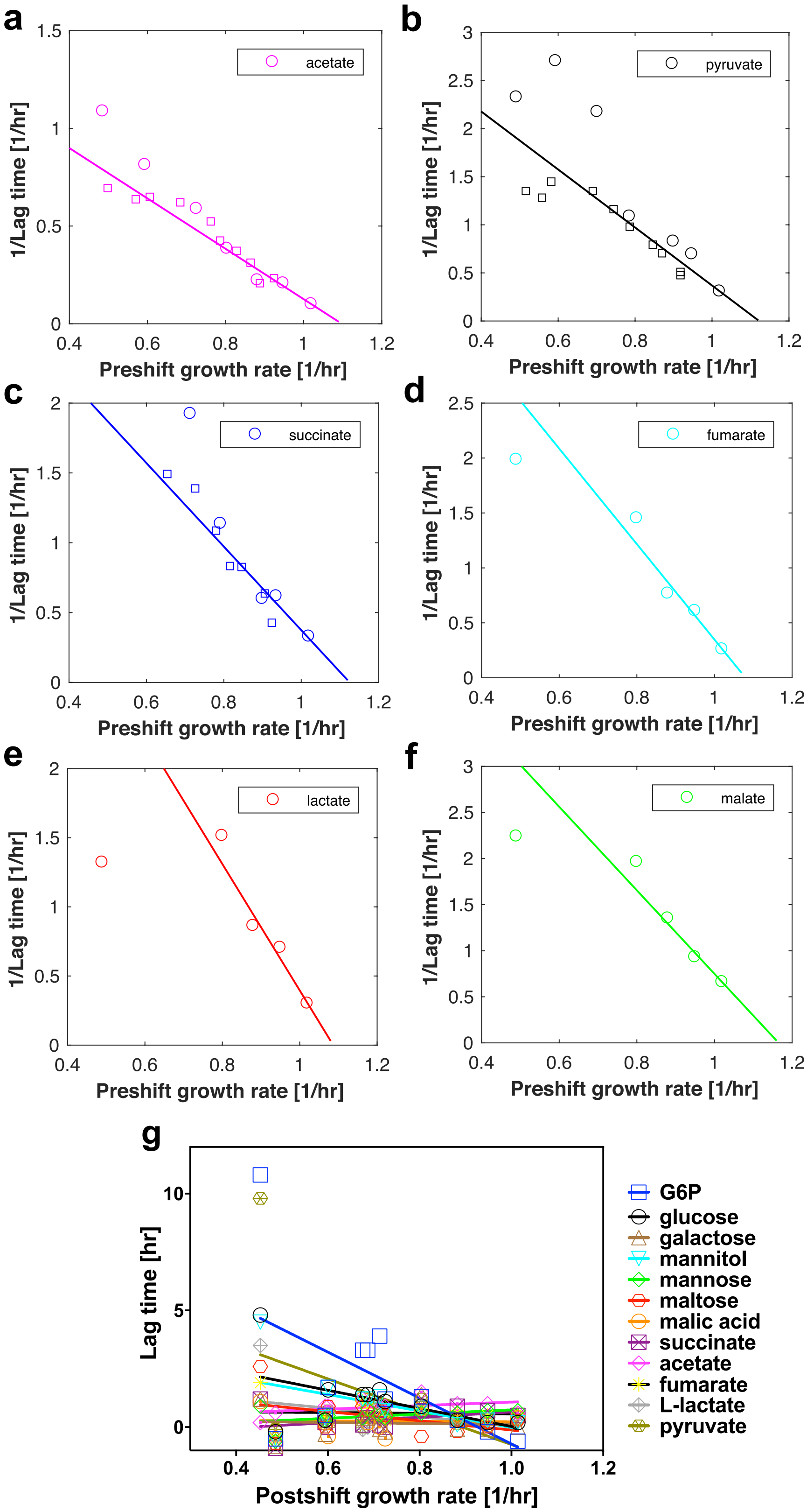
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2505-4.

Correspondence and requests for materials should be addressed to M.B., T.H. or U.S.

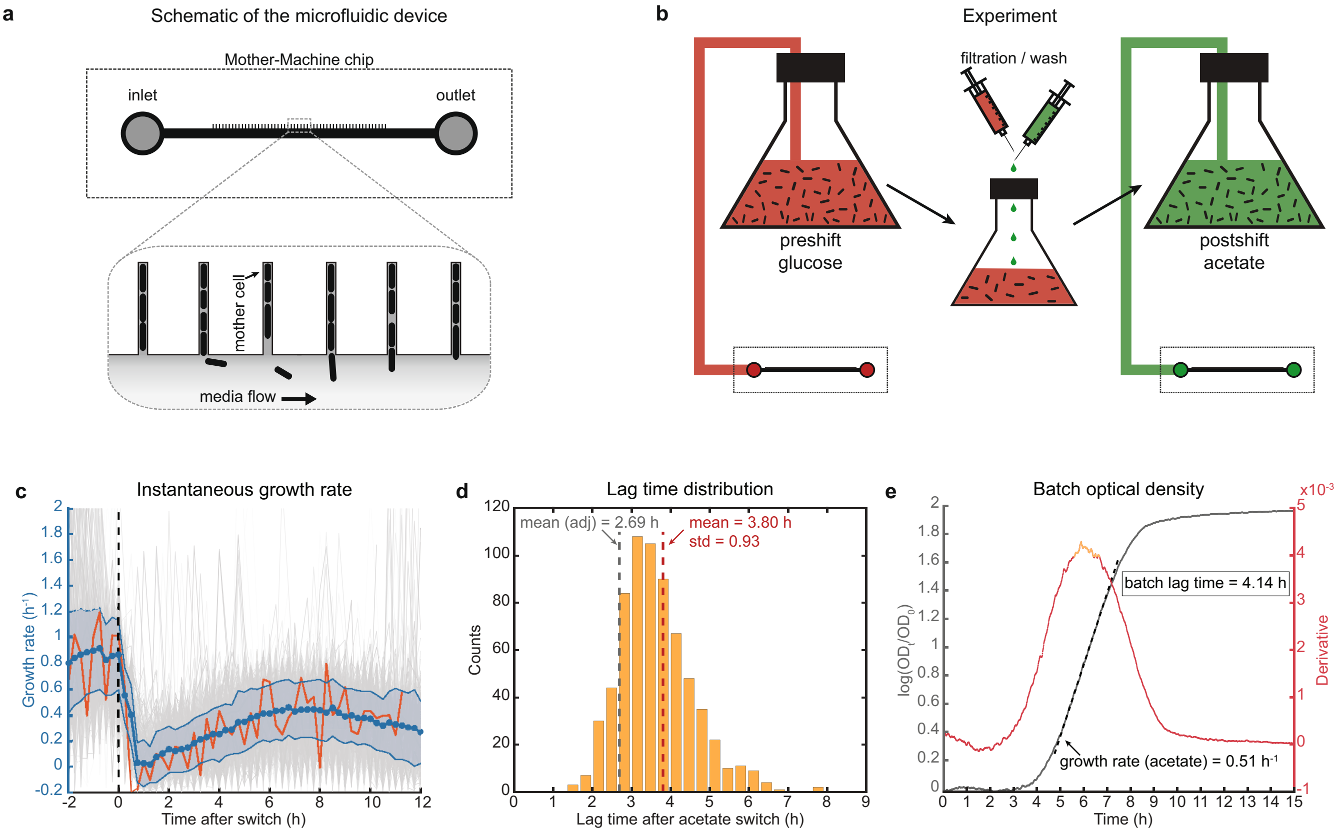
Peer review information *Nature* thanks Jeff Gore, Christopher Marx, Arjan de Visser and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



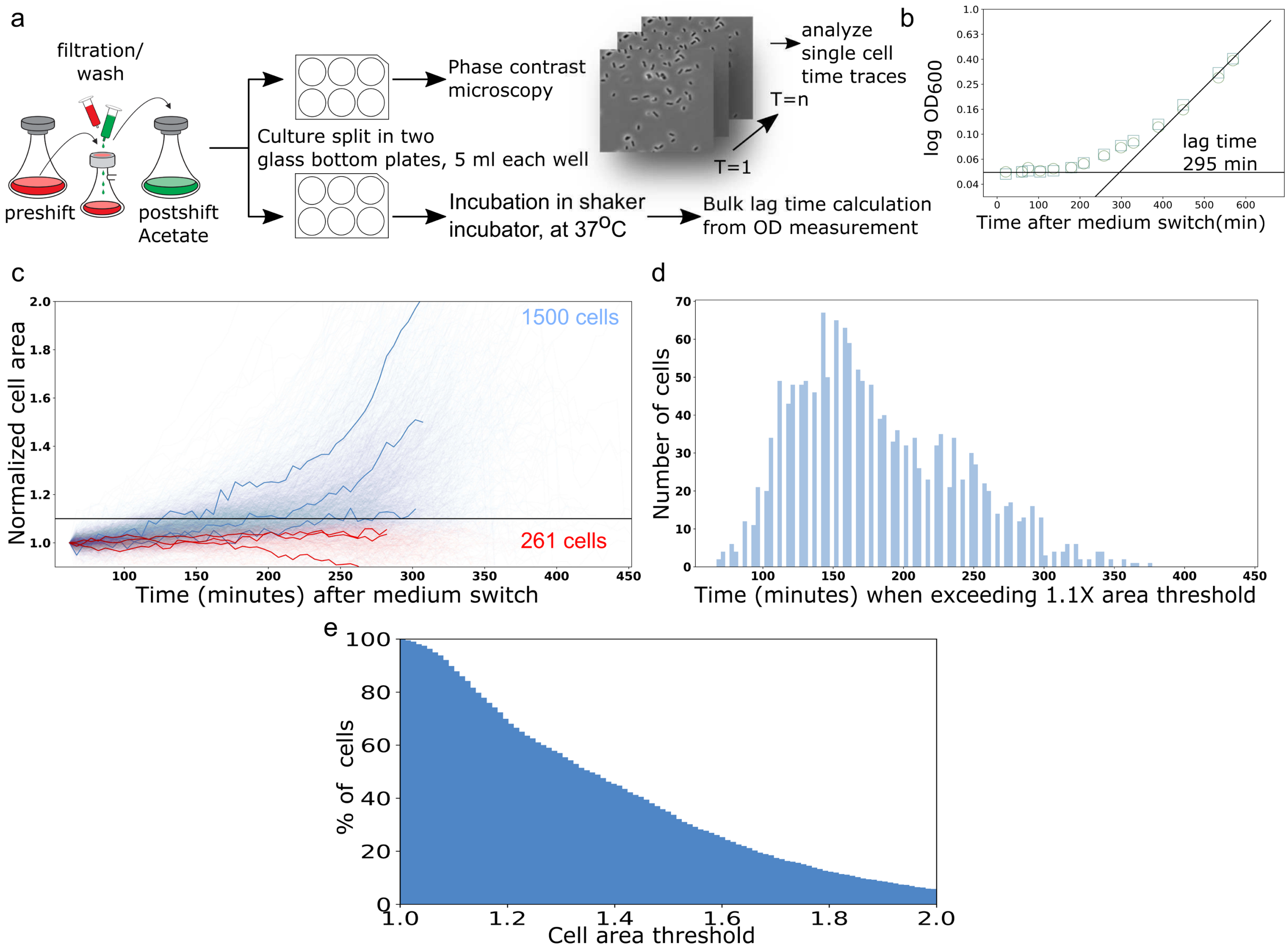
Extended Data Fig. 1 | Growth curves for shifts. **a**, Growth curves following shifts from different glycolytic carbons to acetate by filtration. Long lag phases can consist of several hours without detectable biomass production. There are large variations in the duration of lag phases following shifts from different carbon sources. The duration of the lag phase correlates with the preshift growth rate (Fig. 1): fast growth before the shift results in very long lag times. **b**–**d**, Comparisons of lag times following filtration shifts and in diauxie experiments (which involve no shift, but rather growth on medium containing two sugars, with one sugar running out). **b**, Shift from 1.7 mM glucose to 60 mM acetate. Here the diauxie medium contained glucose plus acetate. **c**, Shift from 1.7 mM glucose to 30 mM succinate. **d**, Shift from 1.7 mM glucose to 40 mM pyruvate. Lag times resulting from filtration shifts and from classical diauxie experiments are mostly comparable. In **c**, the presence of pyruvate in the medium in addition to glucose adversely affected the growth rate, resulting in a shorter lag time in the diauxie shift, consistent with our general observation of the growth-rate dependence of lag times.



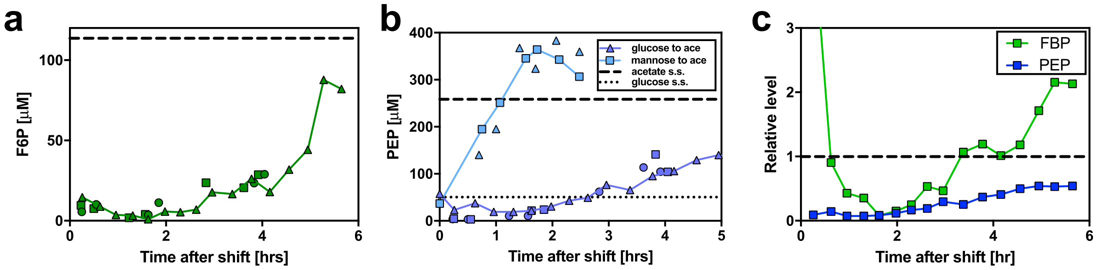
Extended Data Fig. 2 | Lag-time/growth-rate relations. **a**–**f**, The inverse of lag times following a shift to the indicated sugars (**a**, to acetate; **b**, to pyruvate; **c**, to succinate; **d**, to fumarate; **e**, to lactate; **f**, to malate) is plotted as a function of the preshift growth rate in glycolytic conditions. The preshift growth rate was modulated using different carbon sources (circles) and through lactose-uptake titration (squares). Solid lines show nonlinear least-squares fits (Matlab lsqcurvefit function) of lag times as a function of preshift growth rates according to the relation given by equation (1). Most lag phases agree very well with equation (1); only a few shifts, with short lag times (low growth rates), deviate somewhat from this relation. This is partly the result of plotting inverse lag times, which amplifies relatively small experimental variations in lag times for short lag phases. These fits allow us to estimate 95% confidence intervals for model parameters (Matlab nlparci function), most importantly for the critical growth rates λ_0. For acetate, $\lambda_C = (1.10 \pm 0.01)\ h^{-1}$, $\alpha = 0.78 \pm 0.10$, $n = 17$; for pyruvate, $\lambda_C = (1.12 \pm 0.03)\ h^{-1}$, $\alpha = 0.33 \pm 0.07$, $n = 17$; for succinate, $\lambda_C = (1.13 \pm 0.04)\ h^{-1}$, $\alpha = 0.33 \pm 0.10$, $n = 14$; for fumarate, $\lambda_C = (1.08 \pm 0.02)\ h^{-1}$, $\alpha = 0.23 \pm 0.07$, $n = 5$; for lactate, $\lambda_C = (1.09 \pm 0.05)\ h^{-1}$, $\alpha = 0.22 \pm 0.15$, $n = 5$; for malate, $\lambda_C = (1.17 \pm 0.09)\ h^{-1}$, $\alpha = 0.22 \pm 0.11$, $n = 5$. **g**, Lag times as a function of steady-state growth rates in the postshift medium for different preshift media. Coloured solid lines show linear regressions of the corresponding coloured data points. Carbon sources that allow slower growth rates tend to result in longer lag phases when they are the postshift carbon sources. This intuitive correlation has previously been characterized[13].



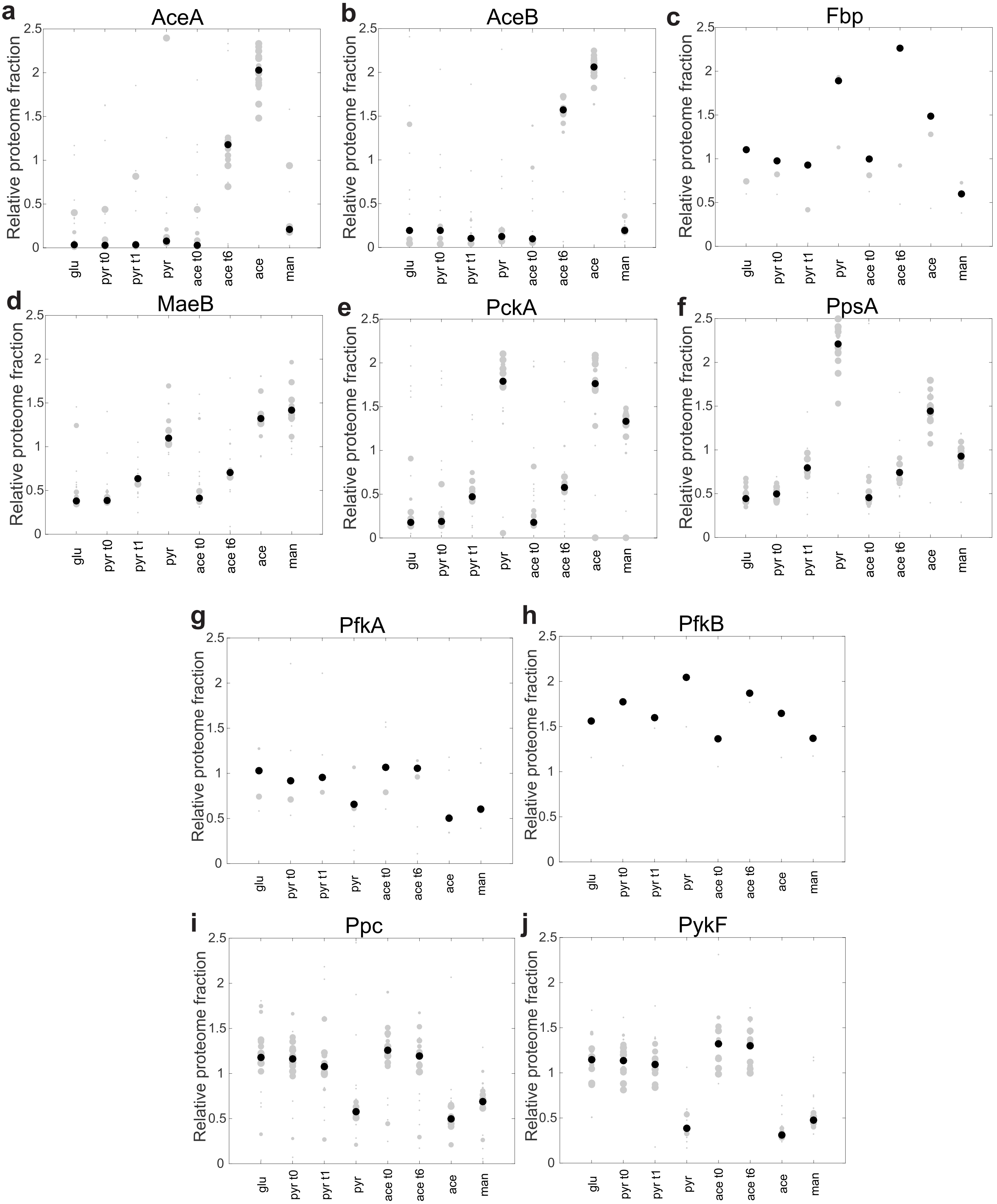
Extended Data Fig. 3 | Single-cell behaviour during a glucose-to-acetate shift in microfluidics. a, Diagram showing the microfluidic device (mother machine) in which bacterial cells are grown. The cells are loaded in narrow trenches (inset), where they are diffusively fed from the medium flowing through the feeding lane. As cells grow out of the trenches, they are washed away by the medium flow. We focused solely on the cells at the closed end of each trench, also called 'mother cells', as they are kept for the duration of the experiment. **b**, Diagram outlining the experimental protocol. Cells were recovered from the mother machine using glucose medium, and then connected to a flask with culture growing in the same medium[43]. The medium was switched as for batch cultures, and the flow was restarted towards the mother machine. Cells continued growing for a short time after filtration both in batch and in the mother machine, presumably because of residual glucose in the system; therefore the experiment resembles a diauxic shift. **c**, Instantaneous single-cell growth rates determined from cell length. Length traces from individual cells were used to compute instantaneous growth rates; the blue points and blue shaded area represent population averages and standard deviations. The orange trace is the instantaneous growth rate trace of an example cell. **d**, Single-cell lag-time distribution. The lag time is defined as the time delay in growth after the switch compared with instantaneous growth at the maximum postshift growth rate. Instantaneous growth-rate traces were used to compute single-cell lag times (Methods). The red dashed line shows the mean of the lag-time distribution of the tracked cells. Cells tracked in the mother machine introduce a bias towards long lag times, because growing cells are washed away instead of being amplified, as happens in batch culture. Therefore, we also calculated the expected batch lag time (2.69 h; grey dashed line), taking into account cell growth (Methods). **e**, The postshift growth curve (grey) of the batch culture connected to the microfluidic chamber was used to determine the batch lag time (4.14 h). The maximum growth rate along the growth curve corresponds to the approximately linear part of the log(OD(*t*)) (grey dotted line), for a growth rate of 0.51 h^{-1}. The red dotted line shows the time derivative of log(OD(*t*)). The quantitative agreement between the microfluidics and the batch is not perfect. Nevertheless, the single-cell distribution of lag times shows that the response of individual cells after the shift is unimodal, and that the lag time is not governed by a small subpopulation of cells that grows immediately on acetate, as expected in ref. [12]. We see no reason why this cell population should not be present in the microfluidics if it were present in the batch. Our data also showed no evidence for the prediction[12] that most cells would never recover and grow after the shift. However, because the cells were grown in a microfluidic chip, our experiment cannot definitively rule out the possibility that the recovery of growth observed here is due to differences in the conditions. To determine whether such a nongrowing population exists in the batch culture, we performed another experiment (Extended Data Fig. 4). *n* = 681 cells. We carried out the growth-curve experiment once, with two independent lanes (one with YFE44, one with the wild-type strain); the plots are relative to results obtained from the flask inoculated with YFE44.



Extended Data Fig. 4 | Single-cell behaviour during a glucose-to-acetate shift through time-lapse microscopy of batch culture. a, Diagram showing the experimental protocol (Methods section on'Batch microscopy'). After the medium shift from glucose to acetate, the culture was split into two identical six-well glass-bottom plates. One was briefly centrifuged and placed into an incubator on a microscope for time-lapse microscopy, and phase-contrast images were recorded. The other plate was placed in a shaker incubator as a control, and OD_{600} was monitored manually. **b**, Growth curves from two biological repeats (circles and squares), obtained by monitoring OD_{600} from the control six-well plate after the media switch. The calculated lag time is 295 min, virtually identical to the batch-culture lag time that we characterized in Fig. 1, indicating that the environment of the six-well plate is almost identical to that of the batch culture as far as the lag time is concerned. **c**, Normalized single-cell-area traces (two biological repeats) from the other plate (n = 1,761 traces). We use cell area as a metric for biomass growth. Light blue traces show the 1,500 cells that crossed an arbitrary 10% threshold for increase in area within our observation time (Methods). Red traces indicate the 261 cells that did not cross this threshold before they became unobservable, either because they detached from the glass or because they were were flooded by other cells. **d**, Histogram showing the distribution of time required for individual cells to increase their area by 10%. **e**, The percentages of cells that grew in cell area (y-axis) by at least the amount shown on the x-axis, relative to their initial size, are plotted. All of these data show that most cells recover after an initial lag phase, eventually growing on acetate. Despite the relatively short observation window of 5–6 h (beyond which the plate became too crowded by dividing cells to allow imaging of individual cells), which is roughly equal to the batch lag time, most cells exhibit substantial growth (**e**). A 10% increase in cell area is easily detectable, and roughly 85% of cells crossed this threshold. The cells that did cross this threshold grew continuously throughout the observation period, exhibiting a single-cell growth curve and a lag time (**c**, **d**) that was similar to the batch lag time. Thus, no more than roughly 15% of cells were completely growth arrested after the shift to actetate, even during this limited observation window. Therefore, in the lag phases studied here, the dormant subpopulations proposed previously[12] had a negligible role in determining lag times. (As an example, even if the roughly 15% of growth-arrested cells never grew again, they would contribute only approximately 21 min to the total lag time of 295 min.).

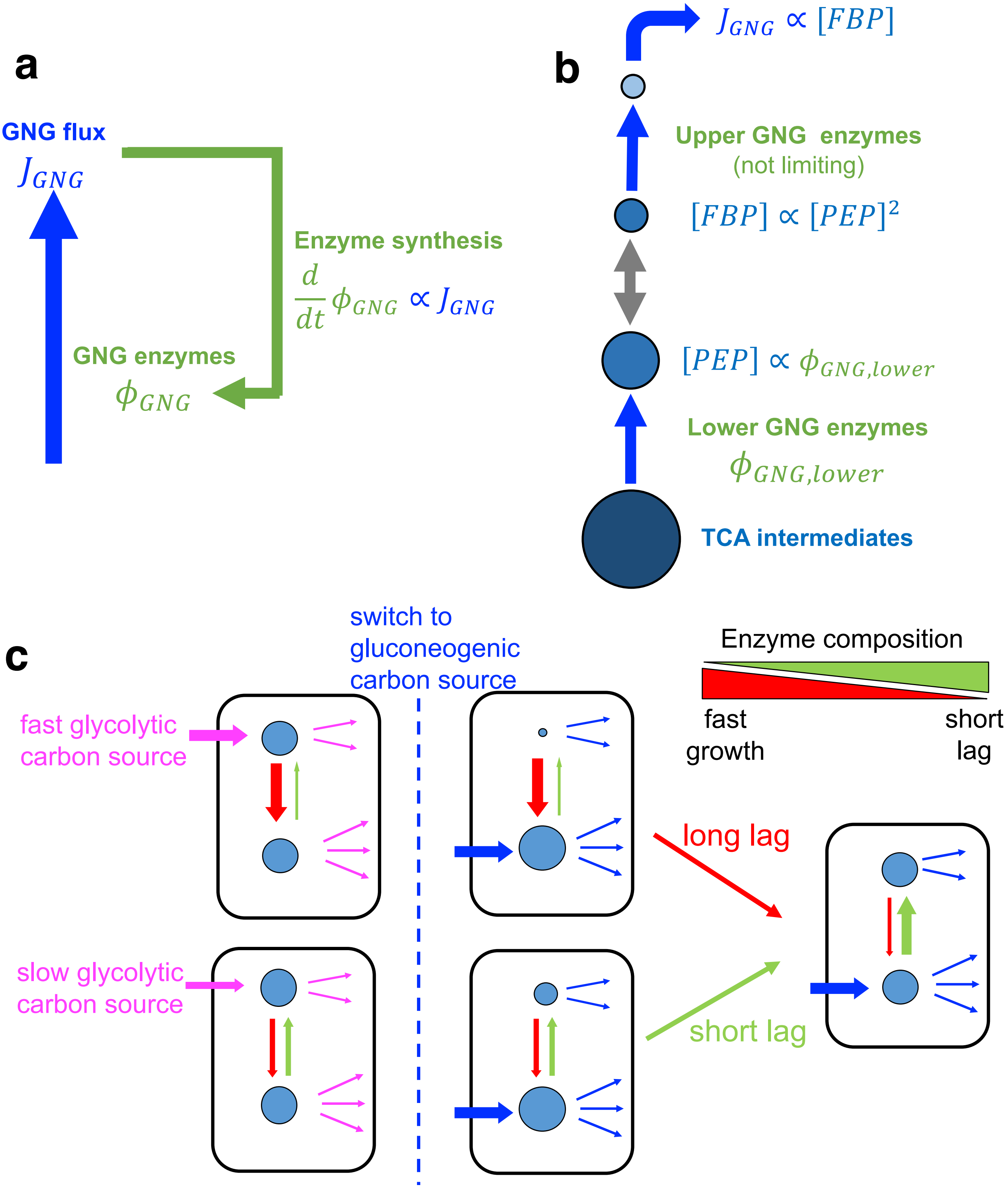


Extended Data Fig. 5 | Absolute and relative concentrations of key metabolites in the shift from glucose to acetate. a, Intracellular concentrations of F6P in the three biological repeats (circles, squares and triangles) of the shift from glucose to acetate presented in Fig. 2. The dashed line represents the steady-state level of F6P for growth on acetate. The F6P concentration is low compared with the Michaelis constants of key enzymes Pgi and TktA, which catalyse the first reactions from F6P in the synthesis of E4P and R5P, essential precursors for biomass production. **b**, Intracellular concentrations of PEP during the lag phase that follows a shift from glucose to acetate and from mannose to acetate. Steady-state (s.s.) concentrations are also shown. PEP acts as a key repressor of glycolytic flux by inhibiting Pfk[51]. The PEP concentration remained low throughout lag phase, even by comparison with the steady-state concentration on glucose, when Pfk is very active. **c**, Time courses of FBP and PEP concentrations throughout lag phase during a shift from glucose to acetate. We normalized the concentrations by their steady-state concentration (dashed line) during exponential growth on acetate. During the lag phase, FBP drops from its steady-state level for growth on glucose, which is more than 100-fold higher than its steady-state level on acetate (normalized to 1). PEP remains at very low concentrations and slowly builds up, together with FBP, 1.5 h after the shift. In our model, we attribute this slow build-up to the need for protein synthesis to increase levels of gluconeogenic enzymes.



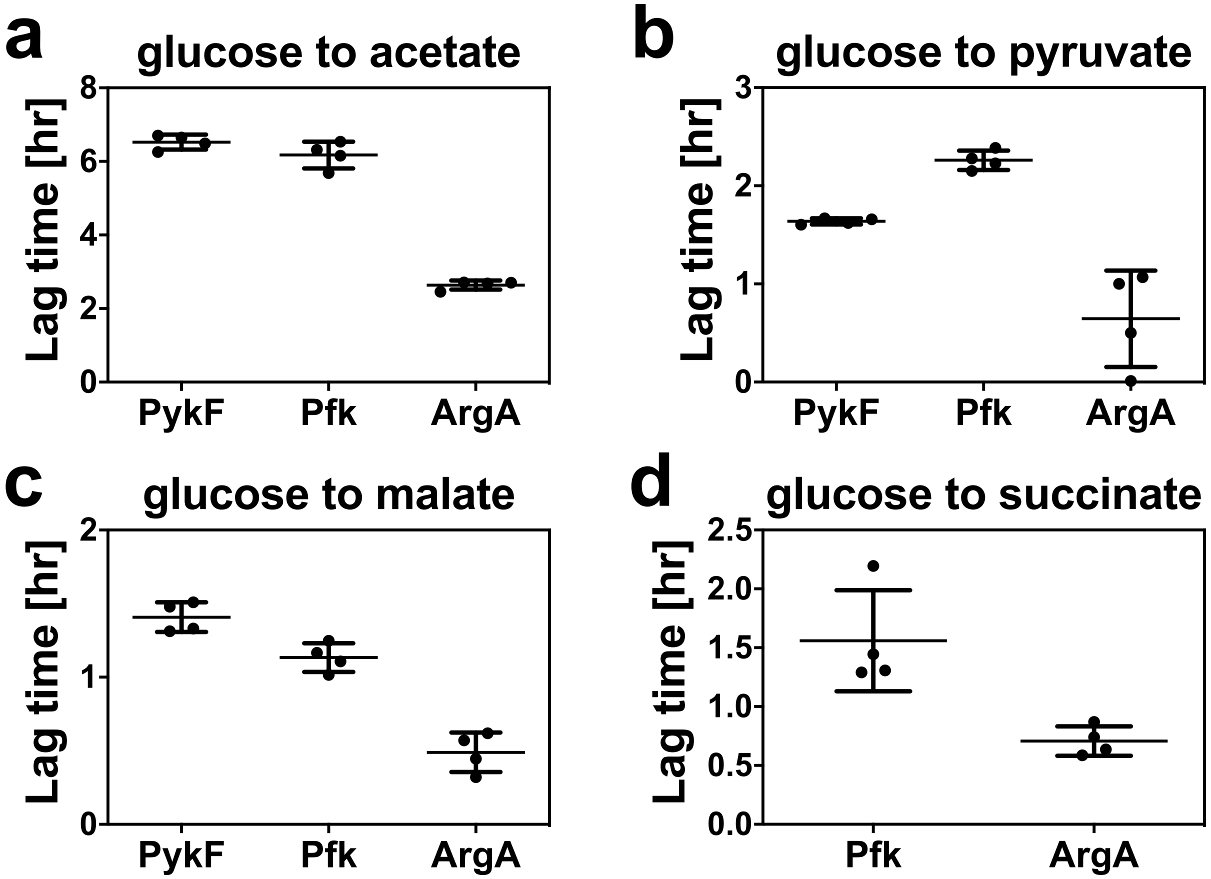
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Proteomics-based characterization of lag-phase dynamics. a–**f**, Gluconeogenic enzymes. Relative levels of gluconeogenic enzymes at different times during lag phase following a shift from glucose to acetate (ace t0, immediately after shift; ace t6, exiting lag phase, 6 h after shift) and glucose to pyruvate (pyr t0, immediately after shift; pyr t1, exiting lag phase, 1 h after shift) and in different steady-state conditions on glucose (glu), pyruvate (pyr), acetate (ace) and mannose (man). **a**, Isocitrate lyase (AceA); **b**, malate synthase (AceB); **c**, fructose-1,6-bisphosphatase (Fbp); **d**, malate dehydrogenase (MaeB); **e**, phosphoenolpyruvate carboxykinase (Pck); **f**, PEP synthase (Pps). **g**–**j**, Glycolytic enzymes. Relative levels of irreversible glycolytic enzymes at different times during lag phase following a shift from glucose to acetate and from glucose to pyruvate, as well as in different steady-state conditions, as for **a**–**f**. **g**, 6-Phosphofructokinase I (PfkA); **h**, 6-phosphofructokinase II (PfkB); **i**, PEP carboxylase (Ppc); **j**, pyruvate kinase I (PykF). Black dots indicate weighted median values. These were obtained from multiple measurements and weighted by the confidence of a sample's quality, as derived from a support vector model (Methods) set up to classify samples into 'high' or 'low' quality, based on a training set of several thousand hand-classified samples[15]. The weights' range is [0,1] and can be found as a separate attribute ('svmPred') for each sample in the accompanying source file. Grey dots indicate individual measurements; the size of each dot indicates the associated confidence (the larger the dot, the higher the confidence that a measurement is of high quality). Dot sizes were defined using the 'MarkerSize' attribute of the 'plot' function in Matlab. Specifically, a dot size was calculated as the confidence value of a measurement (the 'svmPred' attribute) multiplied by 11 (which allowed clearer plotting and ease of visual inspection). If the product of this multiplication for a certain measurement was below a certain minimum value (in our case, 1.8), we set the dot size to this minimum (below that value, the dot would not be visible with the naked eye).

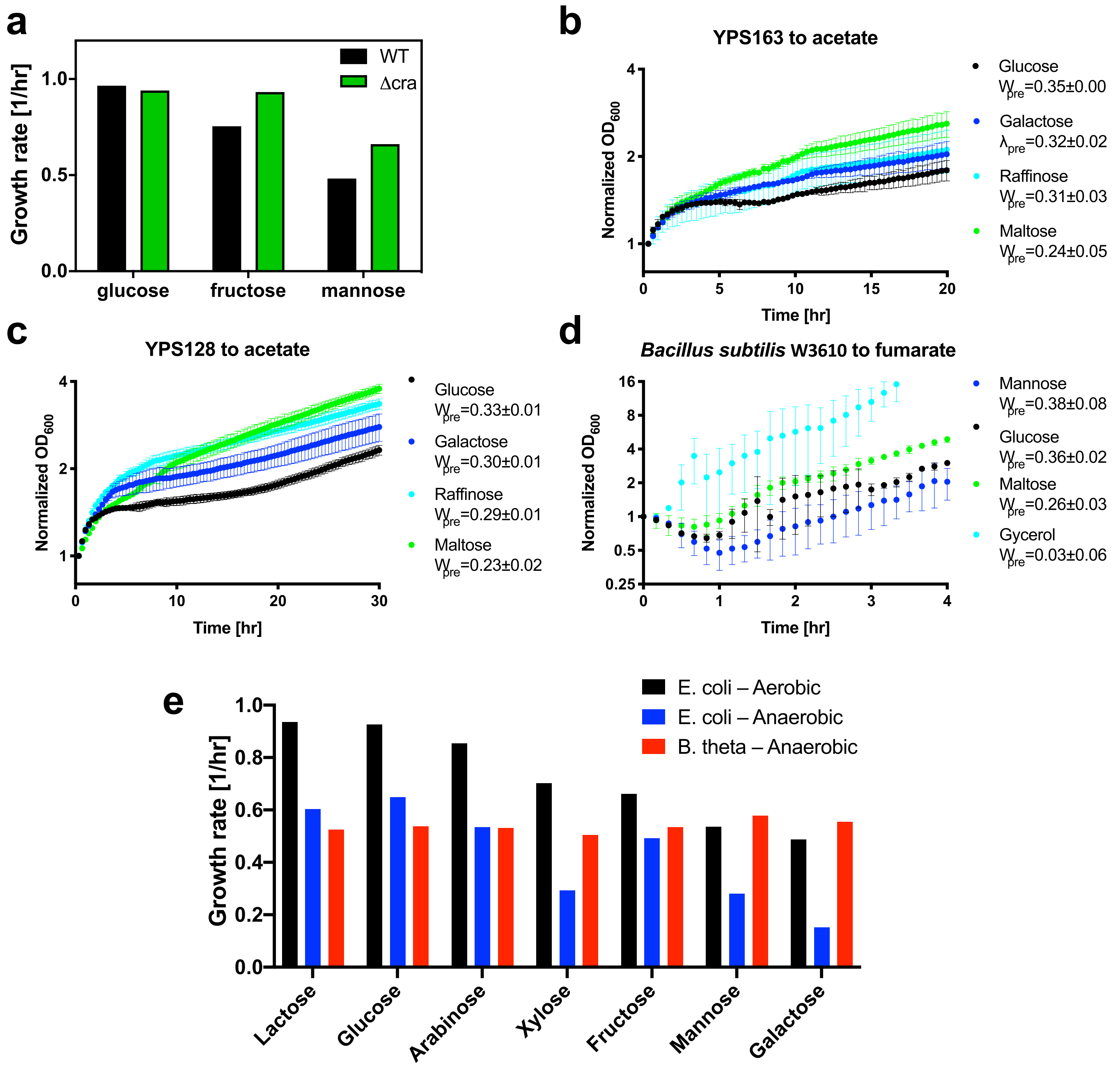


Extended Data Fig. 7 | See next page for caption.

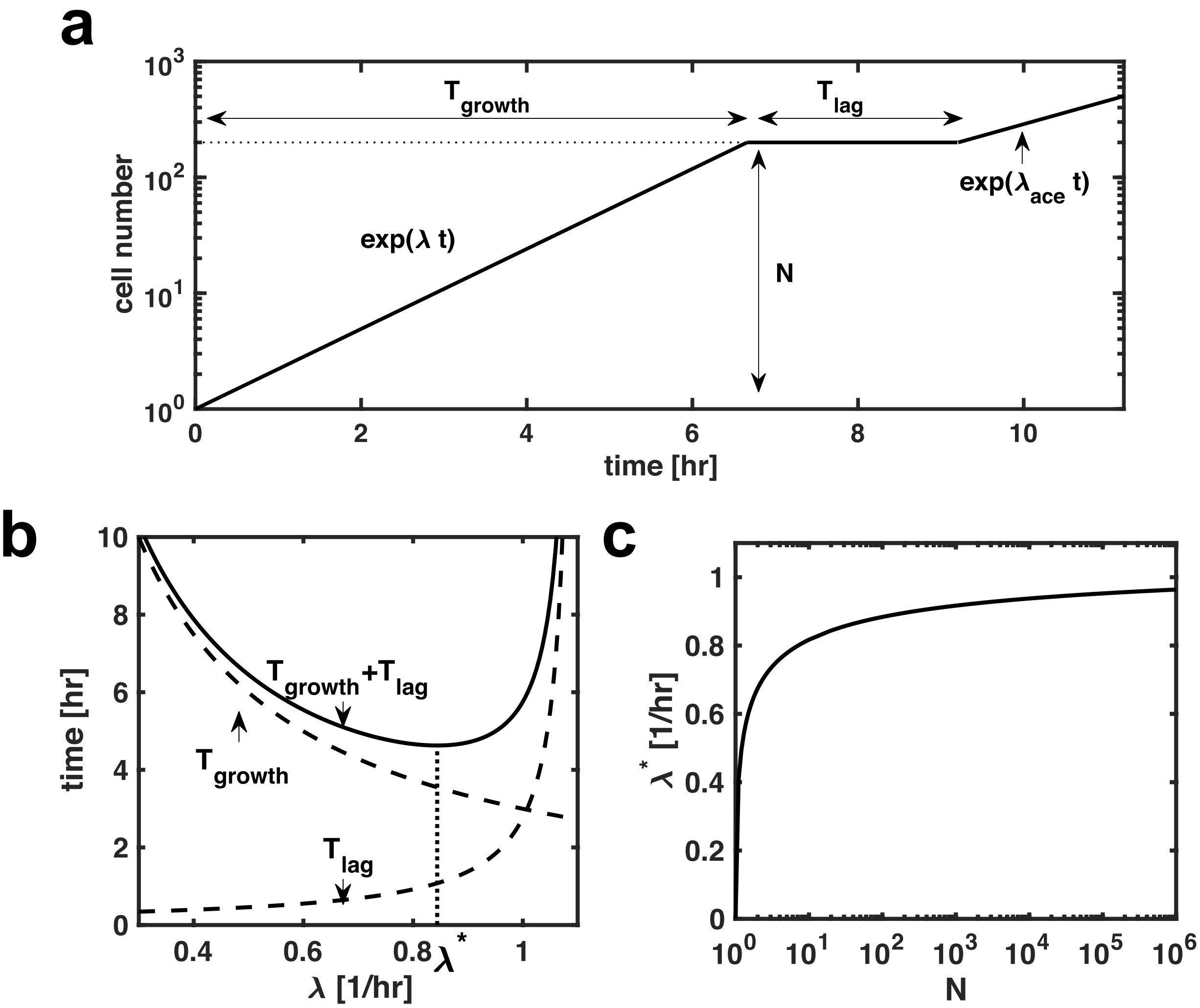
Extended Data Fig. 7 | Sequential flux limitation model and trade-off between growth and lag. a, Intuitively, in our model, lag phases emerge because the gluconeogenic flux, J_{GNG} (blue arrow), limits the synthesis of proteins, which include gluconeogenic enzymes (green arrow). Therefore, the production rate of limiting gluconeogenesis is proportional to the gluconeogenic flux: $\frac{d}{dt}\phi_{\mathrm{GNG,lower}} \propto J_{\mathrm{GNG}}$, in which $\phi_{\mathrm{GNG,lower}}$ denotes the abundance of lower gluconeogenic enzymes. J_{GNG} in turn depends on limiting metabolite concentrations. **b**, To understand the dynamic scaling of these metabolite concentrations, based on the biochemistry of the pathway, we describe gluconeogenesis by a coarse-grained model comprising two irreversible steps (upper and lower gluconeogenesis), connected by reversible reactions. Upper gluconeogenesis does not appear to be limited by its enzyme (Fbp), whose abundance changed only moderately throughout the lag phase and across growth conditions (Extended Data Fig. 6 and proteomics data in ref. [3]). We thus assume the flux through upper gluconeogenesis (top blue arrows) to be limited by the concentration of its substrate, FBP, thus $J_{\mathrm{GNG}} \propto [\mathrm{FBP}]$. The FBP concentration is connected to the output of lower gluconeogenesis, PEP, by the relation $[FBP] \propto [PEP]^2$, owing to the stoichiometry of the reversible reactions (grey arrows). The enzymes of lower gluconeogenesis do appear to be limiting, given previous proteomics data[3] (Fig. 3a and Extended Data Fig. 6). We assume that the lag phase is dominated by a quasistationary period, where transcriptional regulation can be considered constant. The abundances of gluconeogenic enzymes are assumed to change in proportion to each other, characterized by $\phi_{\mathrm{GNG,lower}}$. The latter assumption is plausible, as the expression of gluconeogenic enzymes is primarily controlled by a common transcription factor Cra. Indeed we note that for different preshift (steady-state) conditions, the abundances of different gluconeogenic enzymes are proportional to each other, as they show the same linear growth-rate dependence (Fig. 3a). The flux through lower gluconeogenesis (bottom blue arrow), which is proportional to [PEP], is then governed by $\phi_{\mathrm{GNG,lower}}$. Thus, $[\mathrm{PEP}] \propto \phi_{\mathrm{GNG,lower}}$, resulting in $J_{\mathrm{GNG}} \propto \phi^2_{\mathrm{GNG,lower}}$. **c**, During fast glycolytic growth (top), glycolytic enzymes are highly abundant (thick red arrows), whereas gluconeogenic enzymes are scarce (thin green arrows). The enzyme composition therefore strongly favours glycolysis, resulting in severe depletion of carbon-based metabolites (blue circles) after a shift to gluconeogenic conditions, and hence a long lag phase. For slow glycolytic growth (bottom), the ratio of glycolytic and gluconeogenic enzymes is much more balanced (red and green arrows of similar thickness), resulting in an improved carbon supply to gluconeogenesis after shift and hence a shorter lag. The thick blue and pink arrows illustrate influx from uptake of glycolytic and gluconeogenic substrates, respectively. The thin blue and pink arrows illustrate flux branching off from central carbon metabolism to provide biomass building blocks.



Extended Data Fig. 8 | Preshift overexpression of glycolytic enzymes. a–d, Lag times following shifts from glucose to: **a**, acetate; **b**, pyruvate; **c**, malate; **d**, succinate. The graphs compare the effects of preshift overexpression of the glycolytic enzymes PykF (strain NQ1543) and Pfk (strain NQ1544) with a control enzyme, ArgA (strain NQ1545). Each protein was overexpressed from the same plasmid (pNT3) using the *tac* promoter. Horizontal lines and error bars indicate means and standard deviation ($n = 4$). Lag times more than doubled as a result of preshift overexpression of Pfk or PykF. Thus, the residual activity of glycolytic enzymes is important in lag phase, despite the allosteric regulation of these glycolytic enzymes. Consistent with this picture, the concentration of PEP—a key regulatory metabolite and repressor of glycolytic flux—remained low throughout lag phase, even compared with steady-state levels on glycolytic carbons (Extended Data Fig. 5).



Extended Data Fig. 9 | Improved growth of Cra-knockout *E. coli*, and trade-offs for other microbes. **a**, Growth rates of the Cra knockout (Δcra) on glycolytic carbon sources: growth rates on the slow glycolytic sources (fructose and mannose) are markedly improved compared with the wild type (WT). The Cra knockout expresses very low levels of most gluconeogenic enzymes, and glycolytic enzymes are derepressed; hence it cannot grow on most gluconeogenic carbon sources. **b**–**d**, Growth–adaptation tradeoff in wild-type yeast and *B. subtilis*. We grew two different wild-type yeast strains (YPS163 and YPS128) and a *B. subtilis* strain at different preshift growth rates (λ_{pre}) on different media, before shifting them to acetate (**b**, **c**) and fumarate (**d**) minimal media. After the shift, culture density (OD_{600}) was monitored as a function of time. Data points indicate means; error bars show standard deviations from three biological replicates. The lag time of the growth curves increases with increasing preshift growth, suggesting a trade-off similar to that characterized for *E. coli* (Fig. 1). **e**, Growth comparison for *E. coli* and *B. thetaiotaomicron*, an obligatory anaerobe. The growth rate of *E. coli* NCM3722 on a number of common carbon substrates from the 'top' of central carbon metabolism (glycolysis and pentosephosphate pathways) exhibit a range of values, from 0.9 h^{-1} down to 0.5 h^{-1} (black bars). The growth rates of *B. thetaiotaomicron* (*B. theta*) on the same substrates in anaerobic conditions (red bars) are all within 10% of each other. For comparison, we also show the growth rates of NCM3722 on the same substrates in anaerobic conditions (blue bars). These show a similar pattern of variation as the aerobic growth rates, with the fast ones comparable to that of *B. thetaiotaomicron* (roughly 0.6 h^{-1}) and the slow ones about one-fifth of the fast ones. Saturating amounts of substrates (15 mM) were used, except in the case of *E. coli* on mannose (40 mM).



Extended Data Fig. 10 | Optimal growth rate as a function of the expected substrate abundance in an environment. a, Cells initially grow exponentially by a factor *N* (reflecting the expected carbon abundance) over time T_{growth} at growth rate λ. When carbon runs out, the cells enter lag phase, chartacterized by the lag time, T_{lag}. Cells then again grow exponentially; in the example here, they use the fermentation product acetate at growth rate λ_{ace}. **b**, The optimal strategy for the cell minimizes the total time before postshift exponential growth (resulting in the same cell number, but resuming growth the fastest after the lag phase). The total time before postshift growth resumes is the sum of the growth time, $T_{\mathrm{growth}} = \log(N)/\lambda$, and the lag time, given by equation (1), $T_{\mathrm{lag}} = 1/[\alpha(\lambda_0 - \lambda)]$, both of which are influenced by the growth rate λ. The optimal growth rate, λ^*, minimizes this total time, and is obtained from: $\lambda^* = \lambda_0 \frac{\sqrt{\alpha \ln(N)}}{1+\sqrt{\alpha \ln(N)}}$ **c**, For strain NCM3722, the optimal growth rate, λ^*, given by this equation, is plotted against the expected carbon abundance, given by *N*. The value of α was determined from the fit in Fig. 1d to the majority of glycolytic carbon sources (black line). For realistic carbon abundances, the range of optimal growth rates spans precisely the relatively narrow range of growth rates on naturally occurring carbon sources observed for the wild-type *E. coli* strain NCM3722 (ref. [2]): for example, glucose, 0.95 h^{-1}; mannitol, 0.90 h^{-1}; maltose, 0.79 h^{-1}; glycerol, 0.70 h^{-1}; galactose, 0.59 h^{-1}; mannose, 0.49 h^{-1}. The optimal growth rate drops substantially below 0.5 h^{-1} only when the expected preshift carbon abundance allows for less than a single doubling, $N < 2$, and surpasses 1.0 h^{-1} at enormous, unrealistically high carbon abundances, $N > 10^{12}$, explaining the absence of naturally occurring carbon sources that result in such growth rates.

nature research

Corresponding author(s): Markus Basan
Terence Hwa
Uwe Sauer

Last updated by author(s): Apr 10, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, seeAuthors & Referees and theEditorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	n/a
Data analysis	GraphPad Prism v 8.4.1 MATLAB R2016b

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data for all figures is publicly available. Lag times are provided in Table S2-S3 in the Supplementary Information. All other data is found in downloadable Excel files for each figure. Data from Fig. 3a was taken from Hui et al. Ref. 3 and is deposited on the journal website.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	As many biological repeats as possible were performed. Number of repeats differs from experiment to experiment.
Data exclusions	Not applicable.
Replication	As many biological repeats as possible were performed.
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Structure of the ER membrane complex, a transmembrane-domain insertase

https://doi.org/10.1038/s41586-020-2389-3

Received: 4 February 2019

Accepted: 7 April 2020

Lin Bai[1]✉, Qinglong You[1], Xiang Feng[1], Amanda Kovach[1] & Huilin Li[1]✉

The endoplasmic reticulum (ER) membrane complex (EMC) cooperates with the Sec61 translocon to co-translationally insert a transmembrane helix (TMH) of many multi-pass integral membrane proteins into the ER membrane, and it is also responsible for inserting the TMH of some tail-anchored proteins[1–3]. How EMC accomplishes this feat has been unclear. Here we report the first, to our knowledge, cryo-electron microscopy structure of the eukaryotic EMC. We found that the *Saccharomyces cerevisiae* EMC contains eight subunits (Emc1–6, Emc7 and Emc10), has a large lumenal region and a smaller cytosolic region, and has a transmembrane region formed by Emc4, Emc5 and Emc6 plus the transmembrane domains of Emc1 and Emc3. We identified a five-TMH fold centred around Emc3 that resembles the prokaryotic YidC insertase and that delineates a largely hydrophilic client protein pocket. The transmembrane domain of Emc4 tilts away from the main transmembrane region of EMC and is partially mobile. Mutational studies demonstrated that the flexibility of Emc4 and the hydrophilicity of the client pocket are required for EMC function. The EMC structure reveals notable evolutionary conservation with the prokaryotic insertases[4,5], suggests that eukaryotic TMH insertion involves a similar mechanism, and provides a framework for detailed understanding of membrane insertion for numerous eukaryotic integral membrane proteins and tail-anchored proteins.

Most membrane proteins are synthesized by ribosomes docked on the ER-embedded Sec61 translocon and are folded in the ER membrane. How the topology of so many membrane proteins is maintained is not well understood, but the recently discovered EMC is involved in the process[1–4,6,7].

EMC functions as a TMH insertase for a subset of tail-anchored proteins[3], as well as for the first TMH of many multi-pass integral transmembrane proteins, thereby ensuring their accurate membrane topology in the ER[2]. EMC is also required for the insertion of the second or other TMHs of certain multi-pass integral transmembrane proteins[8–10]. The membrane–protein chaperone function explains why EMC is involved in a diverse set of cellular functions such as protein quality control, biosynthesis of membrane proteins and phospholipids, and virus replication[7,11–14].

The mammalian EMC is composed of 10 subunits, EMC1–EMC10[12] The *Saccharomyces cerevisiae* EMC was first reported to have six subunits, Emc1–Emc6. However, two additional proteins, Sop4 and Ydr056c, were co-purified with Emc1–Emc6[13]. Bioinformatic analysis revealed that the yeast Sop4 and Ydr056c are homologous to the mammalian EMC7 and EMC10, respectively, and therefore, may be the Emc7 and Emc10 subunits of the yeast EMC[15]. To gain a molecular understanding of the activity of EMC, we identified putative EMC substrate or 'client' proteins, purified the endogenous *S. cerevisiae* EMC, determined the cryo-electron microscopy (cryo-EM) structure, and performed functional assays. We found that the yeast EMC is an eight-subunit complex that is evolutionarily conserved with the prokaryotic insertases.

Yeast EMC subunits and client proteins

We inserted a 3× Flag tag onto the carboxyl terminus of the *Emc5* gene in a yeast strain and purified the endogenous EMC by anti-Flag affinity resin and size-exclusion chromatography (Methods, Extended Data Fig. 1a, Supplementary Fig. 1). The SDS–PAGE and mass spectrometry data indicated that the purified EMC complex was composed of eight subunits: Emc1–Emc7 and Emc10 (Fig. 1a). Because the EMC-knockout yeast (missing Emc1–Emc3 and Emc5–Emc6; 5x-Emc) grows normally at 30 °C but has a growth defect at the restrictive temperature of 37 °C[14], we examined the importance of individual Emc subunits for EMC function using the colony growth assay. We found that knocking out any one of the eight subunits led to the same growth defect as the EMC knockout (5x-Emc) at 37 °C (Fig. 1b), which suggests that all subunits are required.

Proteomic analysis of human cells with EMC2, EMC4 or EMC6 knockdown has identified a list of potential EMC client proteins[1,9]. To understand the effect of EMC deficiency and the potential EMC client proteins in yeast, we performed a quantitative proteomic comparison of membrane proteins using tandem mass tag (TMT) labelling in EMC-deficient (Emc3-knockout, Emc4-knockout, or Emc6-knockout) versus wild-type cells. We identified 38 membrane proteins that were significantly reduced; these proteins were likely to be the EMC clients (Fig. 1c, Supplementary Tables 1, 2). We labelled nine selected putative clients with the green fluorescent protein (GFP) reporter and measured their relative membrane abundance in wild-type versus Emc3-knockout

---

[1]Structural Biology Program, Van Andel Institute, Grand Rapids, MI, USA. ✉e-mail: lin.bai@vai.org; huilin.li@vai.org

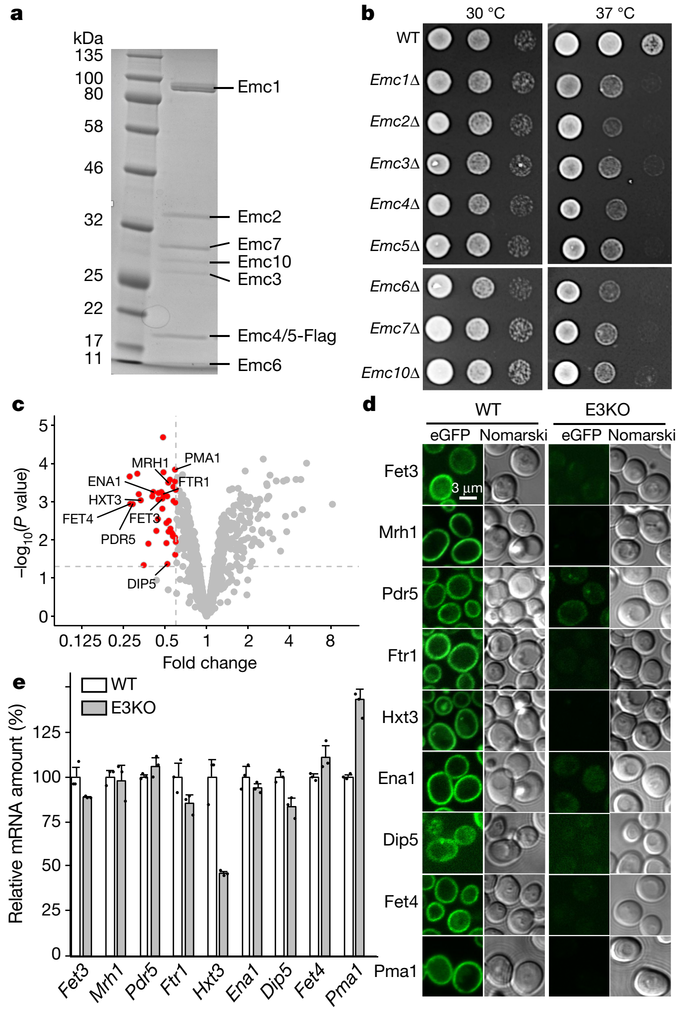


Fig. 1 | Purification of the yeast EMC and identification of EMC client proteins. **a**, The Coomassie blue-stained SDS–PAGE gel of the purified EMC complex. For gel source data, see Supplementary Fig. 1. **b**, Growth of tenfold serial diluted yeast strains (wild-type (WT) and individual Emc subunit knockouts) on YPD plates at 30 °C and 37 °C for 2 days. **c**, Fold change and statistical significance of the membrane protein levels in EMC-knockout relative to wild-type cells. Proteins with a more than 40% decrease in abundance and with $P < 0.05$ are highlighted in red. P values were calculated by empirical Bayes t-tests (two-sided) with no adjustment. **d**, **e**, Protein abundance (**d**) and mRNA levels (**e**) of nine putative EMC clients in wild-type and Emc3-knockout (E3KO) yeast strains. The enhanced GFP (eGFP) reporter is appended to the C termini of the genes. Scale bar, 3 µm. Data are mean ± s.d. Each black dot indicates the value of a single independent experiment. Experiments in **a**–**e** were repeated three times yielding similar results.

yeast cells by fluorescence microscopy (Fig. 1d, Extended Data Fig. 2). The nine proteins were markedly downregulated in Emc3-knockout cells. The downregulation is due to the absence of EMC function rather than transcriptional variation, because the levels of mRNA of these client genes were similar or increased compared to those in the wild-type cells, except for the 50% reduction of *Hxt3* mRNA (Fig. 1e).

Among the 38 putative EMC clients, six (Pdr5, Pdr12, Pho90, Pma1, Ptr2 and Snq2) were found to be associated with EMC[1], and two (Mrh1 and Pma1) were reported to rely on EMC for membrane localization[16,17]. Notably, 16 of the clients had their N termini facing outside; the others faced the cytosol (Supplementary Table 1), which suggests that EMC does not have a preference for the N-terminal location of the client (inside or outside)[1,9].

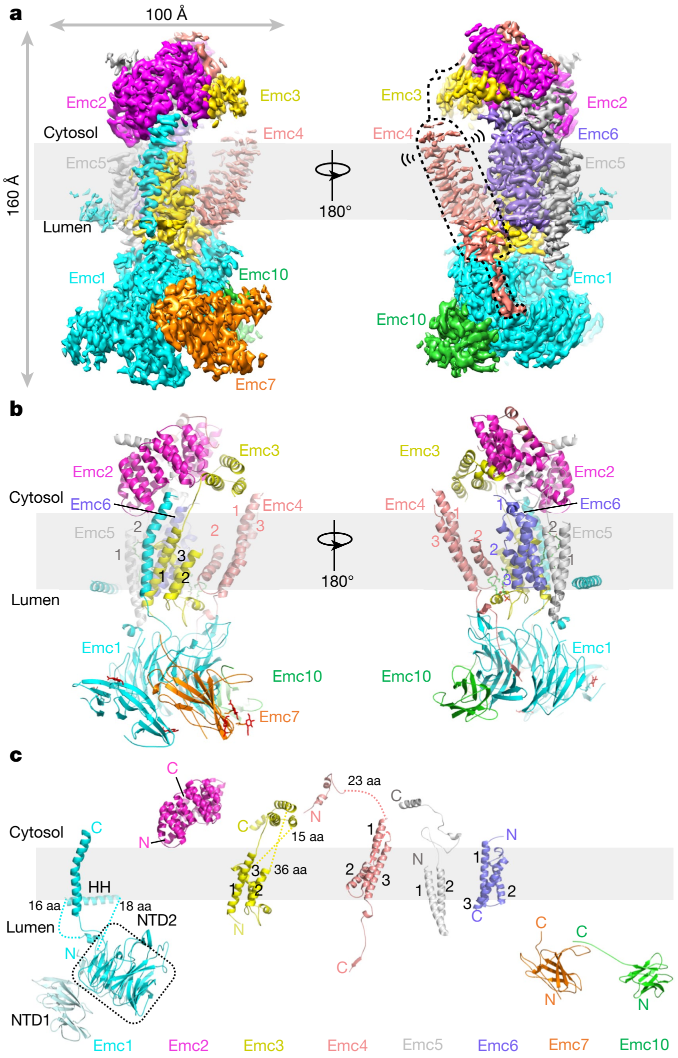


Fig. 2 | Structure of the yeast EMC. **a**, Cryo-EM 3D map of the EMC, showing front and back views with individual subunits coloured. The dotted black shape outlines the Emc4 density, which is weaker and partially flexible (indicated by the two propagating wave signs). **b**, An atomic model shown in cartoon and coloured as in **a**. Phospholipids and N-glycans are shown in green and red, respectively. **c**, Structures of the eight EMC subunits shown separately. aa, amino acids; HH, horizontal α-helix.

EMC architecture and subunit structures

We determined a 3.0-Å average resolution cryo-EM three-dimensional (3D) map (Fig. 2a–c, Extended Data Fig. 1b–g, Extended Data Table 1, Supplementary Video 1). The high resolution allowed us to build the atomic model of EMC de novo (Extended Data Figs. 3, 4, Supplementary Table 3). The structure contained the previously known subunits Emc1–Emc6 plus Emc7 and Emc10 (Figs. 1a, 2a). The EMC structure is approximately 160 × 100 × 80 Å (Fig. 2a, b). Five subunits (Emc1 and Emc3–Emc6) are transmembrane proteins having a total of 12 TMHs. The remaining three subunits—Emc2, Emc7 and Emc10—are aqueous proteins (Fig. 2c). The complex has a transmembrane region, a large lumenal region, and a smaller cytosolic region. There are two ordered phospholipids in the transmembrane region, one facing the lumen and surrounded by Emc3, Emc4 and Emc6, and the other facing the cytosol and surrounded by TMHs of Emc3, Emc5 and Emc6. We identified six N-glycans, three in the lumenal domain of Emc1 (N73, N106 and N192) and three in Emc7 (N53, N85 and N115) (Fig. 2b). We also observed two disulfide bonds, one between C701 and C709 of Emc1 and the other between C65 and C78 of Emc10. The patterns of glycosylation and disulfide bonds were consistent with our membrane orientation assignment of the EMC, in which Emc1, Emc7 and Emc10 are on the lumenal side and Emc2 is in the cytosol. The cytosolic location of Emc2 was supported by the Emc2 interaction with the cytosolic chaperone Hsp90[18]. The EMC cytosolic region is primarily composed of α-helices, whereas the lumenal region is mostly β-strands.

The lumenal region of EMC is formed by Emc1, Emc7 and Emc10 (Fig. 2a, b, Extended Data Fig. 5a–c). The lumenal domain of Emc1 is large and can be further divided into N-terminal domain 1 (NTD1) and NTD2 subdomains (Fig. 2c). The Emc1 NTD2 is an eight-bladed β-propeller, a typical tryptophan–aspartic acid repeat structure (Extended Data Fig. 5a). A structure-based homology search using the online Dali server suggested many homologues, including the fungal ribosomal protein chaperone Sqt1[19], the ribosome assembly protein Rsa4[20], and the ubiquitin ligase SCF complex[21] (Extended Data Fig. 5b). Because these proteins are known to function as a hub to mediate protein–protein or protein–substrate interactions, the structural similarity suggests a similar function for the Emc1 β-propeller. The cytosolic region of EMC is formed by Emc2 and the cytosolic domains of Emc3, Emc4 and Emc5. Emc2 has fifteen α-helices that form seven tetratricopeptide repeats arranged in a right-handed spiral (Extended Data Fig. 5d). The Emc2 tetratricopeptide repeat spiral holds onto the cytosolic regions of Emc3, Emc4 and Emc5 to form the disc-like cytosolic region of EMC that is tilted about 30° away from the ER membrane. EMC was reported to interact with mitochondrial membrane translocase the TOM complex[14]. However, we did not observe direct binding using purified proteins (Extended Data Fig. 6), which suggests that the interaction between EMC and TOM is indirect or too weak to survive the in vitro assay. In the transmembrane region, most TMHs pack tightly against each other except for Emc4 and a horizontal helix of Emc1. The Emc1 horizontal helix is partially embedded in the ER membrane and may stabilize the transmembrane region of EMC (Extended Data Fig. 7). Emc4 has three TMHs that tilt away from Emc3 and Emc6, forming a sizeable cavity in the middle of the complex and creating an opening from the membrane region to the cytosol (Figs. 2a, b, 3a). There is a disordered 23-residue loop at the N-terminal region of Emc4 that enables partial flexibility of the Emc4 TMHs; this dynamism of Emc4 may be relevant to EMC function, as discussed below.

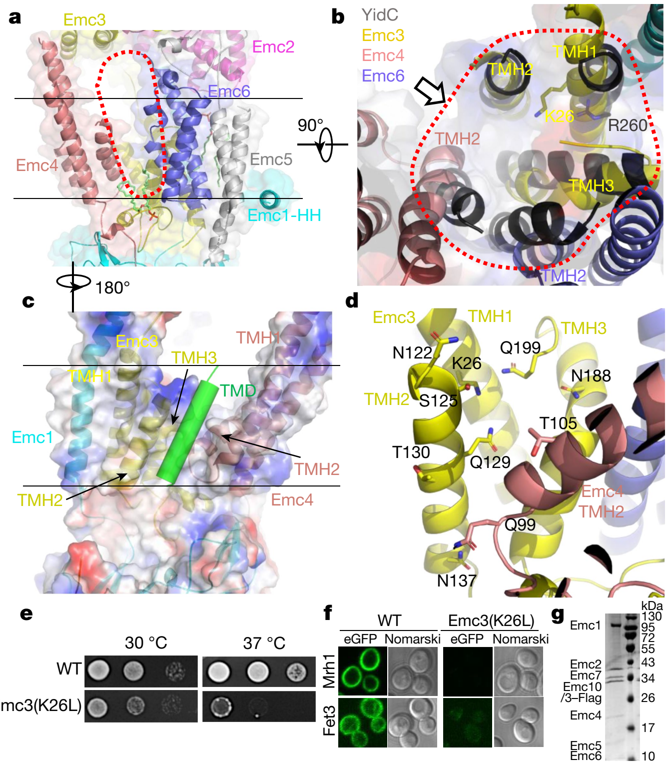


Fig. 3 | The transmembrane region of the yeast EMC contains a client-binding pocket. a, Structure of the transmembrane domain shown as a cartoon in front view. Two parallel black lines mark the lipid bilayer position. The red dots outline the elongated large cavity. Note the horizontal α-helix in Emc1 at the interface between the lumen and the membrane. **b**, Superposition of YidC (PDB code 5Y83) as a black cartoon on the transmembrane domain of EMC in cytosolic view. The red dots encircle the five EMC α-helices aligned with YidC. The putative client TMD position is shown by the arrow, which is suggested by a previous YidC–ribosome EM structure[22]. **c**, A front view of the EMC transmembrane region in cartoon and surface potential. The green cylinder represents a client TMD located between TMH2 of Emc3 and TMH2 of Emc4 in the putative client-binding pocket. Panels **c** and **d** are viewed from the back of panel **a**. **d**, The polar environment of the putative client-binding pocket of the EMC. **e**, Two-day growth of tenfold serially diluted cells (wild type and Emc3(K26L) mutant) on YPD plates at 30 °C and 37 °C. **f**, Diminished amount of two EMC clients (Mrh1 and Fet3) in cells containing the Emc3(K26L) mutation. eGFP is inserted into the C termini of these genes. **g**, The Coomassie blue-stained SDS–PAGE gel of the purified mutant EMC containing a K26L single mutation in Emc3. Experiments in **e**–**g** were repeated three times yielding similar results. For gel source data, see Supplementary Fig. 1.

The substrate-binding pocket in EMC

The EMC transmembrane region contains a large cavity surrounded by Emc3, Emc4 and Emc6, and the cavity is accessible from either the front or the left side in the membrane (Fig. 3a, b). EMC is expected to have a TMH-binding pocket to facilitate insertion of a client TMH into the ER membrane. The cavity inside the transmembrane region is the only site with enough size to accommodate a TMH. A previous bioinformatic analysis identified Emc3 as a member of the evolutionarily conserved Oxa1–Alb3–YidC family, which inserts tail-anchored proteins; that family includes the eukaryotic insertase Get1 and the prokaryotic insertase YidC[5]. In contrast to Emc3 and Get1, which each have three TMHs, YidC has five TMHs (TM2–TM6) and an amphipathic horizontal helix (EH1)[5] (Extended Data Fig. 8a–c).

We found that the three TMHs of Emc3, together with TMH2 of Emc4 and TMH2 of Emc6, form a YidC-like fold (Fig. 3b). These five TMHs of EMC contain a client-binding groove as in the YidC structure. In the EM structures of the YidC–ribosome complexes, the TMH of a nascent peptide emerging from the ribosome is located between TM3 and TM5 in the YidC structure, which corresponds in the EMC to TMH2 of Emc3 and TMH2 of Emc4[22–24]. We suggest that this is the binding site of the EMC client, based on the marked structural conservation between the EMC and YidC (Fig. 3c). Notably, this site is located on the Emc3 side in the central cavity. The surface electrostatic potential around the client site is a mix of charges and hydrophobicity. Many polar residues—including K26, N122, S125, T130, N137, N188, Q129 and Q199 of Emc3, and Q99 and T105 of Emc4—outline the ends of the client site (Fig. 3d). The middle of the client pocket is relatively hydrophobic. It is uncommon to have so many polar residues exposed to the hydrophobic membrane environment, but this feature is consistent with the preference of the EMC for moderately hydrophobic or partially hydrophilic TMHs[2].

The hydrophilic groove in EMC features a positively charged residue (K26 of Emc3), which is structurally equivalent to R72 in the *Bacillus halodurans* YidC[25], R260 in the *Thermotoga maritima* YidC[26] (Fig. 3b, d), and R366 in the *Escherichia coli* YidC[24,27]. The hydrophilicity of the client grooves of the YidC structures is important for substrate binding[26]. This knowledge is consistent with the finding that increasing the hydrophobicity of a client makes it less dependent on the EMC for membrane insertion, and conversely, that increasing the hydrophilicity makes the client more dependent on the EMC[2,9]. We produced a mutant Emc3(K26L) yeast strain and found that the cells grew much slower than wild-type cells at the increased temperature of 37 °C (Fig. 3e). Furthermore, the putative EMC clients (Mrh1 and Fet3) were unable

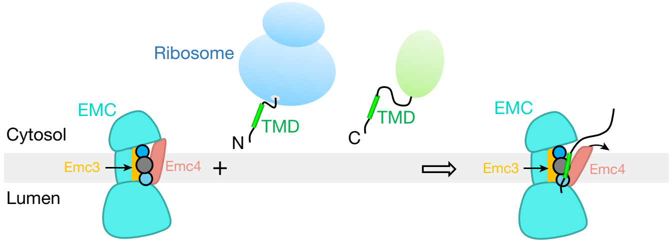


Fig. 4 | A model for client TMH insertion by the eukaryotic EMC. The model highlights the ability of EMC to chaperone or facilitate membrane insertion of a diverse set of transmembrane protein clients, with their respective TMH either at the N terminus or at the C terminus. The TMH insertion can be either co-translational (represented by a client emerging from a ribosome) or post-translational (represented by a client with a folded green domain). The model also shows the presence of a partially hydrophilic pocket formed by the TMDs of Emc3 and Emc4—the putative client-binding pocket—in the transmembrane region of the EMC complex. The pocket is lined by three connected circles, which represent the presence of multiple hydrophilic (blue circles) and hydrophobic residues (grey circle). The curved black arrow indicates a potential movement of the Emc4 TMD to accommodate the client TMH.

to properly fold and locate to the membranes of the Emc3(K26L) cells (Fig. 3f, Extended Data Fig. 9a). We confirmed that the Emc3(K26L) mutation did not affect EMC assembly, because the intact mutant complex could be purified (Fig. 3g). These results support our assignment of the partially hydrophilic cavity as the client-binding site.

In YidC, TMH2 and TMH3 move away from TMH4–TMH6, widening the central groove between TMH3 and TMH5 to accommodate the client TMH[22–24]. The corresponding movement in EMC is between Emc3 TMH2 and Emc4 TMH2. To test whether the flexibility of the transmembrane domains (TMDs) of Emc4 enables a similar conformational change in EMC, we prepared three mutant yeast strains by truncating 5, 10 or 15 residues from the 23-residue loop in Emc4. All strains lost EMC function, as shown by their growth defect at 37 °C (Extended Data Fig. 9b).

EMC resembles YidC in two additional ways: first, both the Emc1 horizontal helix and EH1 of YidC are partially embedded in the exoplasmic side of the membrane to support other TMHs; and second, the lumenal region of EMC and the periplasmic P1 domain of YidC are both primarily composed of β-strands (Extended Data Fig. 8a–c). The EMC lumenal region may also interact with the Sec translocon, similar to the YidC P1 domain[28,29].

A model for client TMH insertion by EMC

EMC inserts tail-anchored proteins and the first TMH of membrane proteins[2,3], as well as the second or other TMHs for some multi-pass integral transmembrane proteins[8–10]. How EMC recognizes such diverse clients is unclear. By combining our studies with recent biochemical work[1,2], we suggest a client TMH insertion mechanism for the EMC as shown in Fig. 4.

A key feature of an EMC client is the partial hydrophilicity of the TMH—that is, it contains several polar or charged residues[2,8,9,30]. To accommodate such clients, the client-binding pocket of EMC is also partially hydrophilic. Emc3 is at the core of the EMC active site, consistent with its evolutionary link with the Oxa1–Alb3–YidC insertase family. Another important feature of EMC is the flexible client-binding pocket, made possible by the long linker connecting the TMD of Emc4. Similar flexibility is also observed in the homologue YidC[22–24].

Therefore, this study reveals a notable structural and mechanistic conservation between the eukaryotic EMC and the prokaryotic insertases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-020-2389-3.

1. Shurtleff, M. J. et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018).
2. Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC is required to initiate accurate membrane protein topogenesis. *Cell* **175**, 1507–1519 (2018).
3. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470–473 (2018).
4. Chen, Y. & Dalbey, R. E. Oxa1 superfamily: new members found in the ER. *Trends Biochem. Sci.* **43**, 151–153 (2018).
5. Anghel, S. A., McGilvray, P. T., Hegde, R. S. & Keenan, R. J. Identification of oxa1 homologs operating in the eukaryotic endoplasmic reticulum. *Cell Rep.* **21**, 3708–3716 (2017).
6. Fry, M. Y. & Clemons, W. M. Jr. Complexity in targeting membrane proteins. *Science* **359**, 390–391 (2018).
7. Richard, M., Boulin, T., Robert, V. J., Richmond, J. E. & Bessereau, J. L. Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proc. Natl Acad. Sci. USA* **110**, E1055–E1063 (2013).
8. Coelho, J. P. L. et al. A network of chaperones prevents and detects failures in membrane protein lipid bilayer integration. *Nat. Commun.* **10**, 672 (2019).
9. Tian, S. et al. Proteomic analysis identifies membrane proteins dependent on the ER membrane protein complex. *Cell Rep.* **28**, 2517–2526 (2019).
10. Hiramatsu, N., Tago, T., Satoh, T. & Satoh, A. K. ER membrane protein complex is required for the insertions of late-synthesized transmembrane helices of Rh1 in *Drosophila* photoreceptors. *Mol. Biol. Cell* **30**, 2859–2942 (2019).
11. Bagchi, P., Inoue, T. & Tsai, B. EMC1-dependent stabilization drives membrane penetration of a partially destabilized non-enveloped virus. *eLife* **5**, e21470 (2016).
12. Christianson, J. C. et al. Defining human ERAD networks through an integrative mapping strategy. *Nat. Cell Biol.* **14**, 93–105 (2011).
13. Jonikas, M. C. et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
14. Lahiri, S. et al. A conserved endoplasmic reticulum membrane protein complex (EMC) facilitates phospholipid transfer from the ER to mitochondria. *PLoS Biol.* **12**, e1001969 (2014).
15. Wideman, J. G. The ubiquitous and ancient ER membrane protein complex (EMC): tether or not? *F1000Res.* **4**, 624 (2015).
16. Luo, W.-J., Gong, X.-H. & Chang, A. An ER membrane protein, Sop4, facilitates ER export of the yeast plasma membrane [H⁺]ATPase, Pma1. *Traffic* **3**, 730–739 (2002).
17. Bircham, P. W. et al. Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. *Mol. Biosyst.* **7**, 2589–2598 (2011).
18. Kudze, T., Mendez-Dorantes, C., Jalloh, C. S. & McClellan, A. J. Evidence for interaction between Hsp90 and the ER membrane complex. *Cell Stress Chaperones* **23**, 1101–1115 (2018).
19. Pausch, P. et al. Co-translational capturing of nascent ribosomal proteins by their dedicated chaperones. *Nat. Commun.* **6**, 7494 (2015).
20. Baßler, J. et al. A network of assembly factors is involved in remodeling rRNA elements during preribosome maturation. *J. Cell Biol.* **210**, 169–170 (2015).
21. Tang, X. et al. Composite low affinity interactions dictate recognition of the cyclin-dependent kinase inhibitor Sic1 by the SCFCdc4 ubiquitin ligase. *Proc. Natl Acad. Sci. USA* **109**, 3287–3292 (2012).
22. Kedrov, A. et al. Structural dynamics of the YidC:ribosome complex during membrane protein biogenesis. *Cell Rep.* **17**, 2943–2954 (2016).
23. Wickles, S. et al. A structural model of the active ribosome-bound membrane protein insertase YidC. *eLife* **3**, e03035 (2014).
24. Kohler, R. et al. YidC and Oxa1 form dimeric insertion pores on the translating ribosome. *Mol. Cell* **34**, 344–353 (2009).
25. Kumazaki, K. et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516–520 (2014).
26. Xin, Y. et al. Structure of YidC from *Thermotoga maritima* and its implications for YidC-mediated membrane protein insertion. *FASEB J.* **32**, 2411–2421 (2018).
27. Kumazaki, K. et al. Crystal structure of *Escherichia coli* YidC, a membrane protein chaperone and insertase. *Sci. Rep.* **4**, 7299 (2014).
28. Xie, K., Kiefer, D., Nagler, G., Dalbey, R. E. & Kuhn, A. Different regions of the nonconserved large periplasmic domain of *Escherichia coli* YidC are involved in the SecF interaction and membrane insertase activity. *Biochemistry* **45**, 13401–13408 (2006).
29. Sachelaru, I. et al. YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. *J. Biol. Chem.* **288**, 16295–16307 (2013).
30. Lin, D. L. et al. The ER membrane protein complex promotes biogenesis of dengue and Zika virus non-structural multi-pass transmembrane proteins to support infection. *Cell Rep.* **27**, 1666–1674 (2019).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Purification of the endogenous EMC complex

The C-terminal, triple-Flag-tagged Emc5 construct was generated by using a PCR-based genomic epitope-tagging method[31] on the yeast strain W303-1a (*MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11*). 18L cells were grown in YPD medium for about 20 h. The collected cells were resuspended in lysis buffer containing 20 mM Tris-HCl (pH 7.4), 0.2 M sorbitol, 50 mM potassium acetate, 2 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride (PMSF) and then were lysed using a French press at 15,000 psi. Lysate was centrifuged at 10,000*g* for 30 min at 4 °C. The supernatant was collected and centrifuged at 100,000*g* for 60 min at 4 °C. The membrane pellet was collected and then resuspended in buffer A containing 10% glycerol, 20 mM Tris-HCl (pH 7.4), 1.5% digitonin, 0.5 M NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM EDTA, and 1 mM PMSF. After incubation for 30 min at 4 °C, the mixture was centrifuged for 30 min at 120,000*g* to remove insolubilized membrane. The supernatant was mixed with pre-washed anti-Flag (M2) affinity gel at 4 °C overnight with shaking. The affinity gel was then collected and washed three times in buffer B containing 0.1% digitonin, 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, 1 mM $MgCl_2$ and 1 mM $MnCl_2$. The EMC was eluted with buffer B containing 0.15 mg ml^{-1} 3× Flag peptide and was further purified in a Superose 6 10/300 gel filtration column in buffer C containing 0.1% digitonin, 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, 1 mM $MgCl_2$ and 1 mM $MnCl_2$. Finally, the purified EMC sample was assessed by SDS–PAGE gel and the subunit composition was identified by trypsin digestion and mass spectrometry.

Colony growth assay

Yeast wild-type (BY4741) and EMC-knockout strains were purchased from The Yeast Knockout (YKO) Collection of Horizon Discovery. Emc3 mutants and truncations—Emc3(K26L), Emc4(Δ56–60), Emc4(Δ51–60) and Emc4(Δ46–60)—were prepared using plasmid pFA6a-His3 in the BY4741 strain. The strains were first grown to the same OD in the YPD medium at 30 °C. Then, 7 μl of 1:10 serial dilutions of the cells were spotted onto YPD plates, incubated at 30 °C or 37 °C for 2 days and then were examined for growth phenotype.

TMT mass spectrometry

The membrane pellets were prepared following the above-described method for EMC purification. Then the membrane preparations were resuspended in buffer containing 10% glycerol, 20 mM Tris-HCl (pH 7.4), 1% DDM, 0.5 M NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM EDTA and 1 mM PMSF. After centrifugation at 100,000*g* for 60 min at 4 °C, the supernatants were collected and sent to MS Biowork for tandem mass tag (TMT) mass spectrometry. Data analysis followed the protocol using scripts published previously (http://www.biostat.jhsph.edu/~kkammers/software/eupa/R_guide.html)[32]. Only proteins that are annotated to be membrane proteins in Gene Ontology annotation were plotted.

Light microscopy and image processing

Genes were labelled by eGFP in the C termini using plasmid pFA6a-link-yoEGFP-SpHis5 (Addgene). Microscopy was performed with a Nikon A1plus-RSi laser scanning confocal microscope at 100× oil objective. Image acquisition and analysis were performed with the program NIS-Elements Software and ImageJ. The displayed microscopic images of control and knockout/mutant samples were adjusted equally using the same brightness and contrast values. Yeast cells were briefly washed with water and immediately imaged in water at room temperature.

RNA isolation and quantitative real-time PCR

Total RNA was extracted from cells with MasterPure Yeast RNA Purification Kit (Lucigen). The SuperScript IV VILO Master Mix Kit (Invitrogen Life Technologies) was used for first-strand complementary DNA synthesis (0.1 μg μl^{-1} mRNA in reaction system). Quantitative PCR amplification was carried out using the Step One Plus Thermocycler (Applied Biosystems). Each reaction included 5 μl Power SYBR Green Real-Time PCR Master Mix (Applied Biosystems), 2.5 μl complementary DNA sample and 2.5 μl PCR primer mix (forward and reverse each 0.8 μm). Actin (*ACT1*) was used as internal control. The relative gene expression was expressed as a percentage of the wild-type control.

Cryo-EM

Aliquots of 3 μl of purified EMC at a concentration of about 5 mg ml^{-1} were placed on glow-discharged holey carbon grids (Quantifoil Au R2/1, 300 mesh) and were flash-frozen in liquid ethane using an FEI Vitrobot Mark IV. Cryo-EM data were collected automatically with SerialEM in a 300-kV FEI Titan Krios electron microscope operated at a nominal magnification of 130,000× and a pixel size of 0.5145 Å per pixel with defocus values from −1.0 to −2.0 μm. A K2 direct detector was used for image recording under counting mode. The dose rate was 8.6 electrons per $Å^2$ per second, and the total exposure time was 8 s. The total dose was divided into a 40-frame movie so each frame was exposed for 0.2 s.

Cryo-EM image processing

We collected 4,260 raw movie micrographs. Program MotionCorr 2.0[33] was used for motion correction, and CTFFIND 4.1 was used for calculating contrast transfer function parameters[34]. All the remaining steps were performed using RELION 3[35]. Templates for automatic particle picking were generated from a two-dimensional (2D) classification of about 2,000 manually picked particles. A total of 590,118 particles were picked automatically. 2D classification was then performed, and particles in the classes with features unrecognizable by visual inspection were removed. A total of 464,190 particles remained and were used for 3D classification. Based on the quality of the four 3D classes, 355,991 particles belonging to two good classes were selected for further 3D reconstruction, refinement, and post-processing, resulting in a 3.0-Å average resolution 3D density map. The resolution of the map was estimated by the gold-standard Fourier shell correlation at a correlation cut-off value of 0.143.

Structural modelling, refinement, and validation

The initial models of EMC were first automatically built into the 3.0-Å EM map using the map_to_model in the PHENIX program[36]. About 1,000 residues (approximately 60% of the whole complex) were automatically modelled, and about half of them were Cαs. The initial model was then manually checked and corrected in COOT[5]. On the basis of the initial model, we then manually built the entire complex in the programs COOT[5] and Chimera[37,38]. The complete EMC model was refined by real-space refinement in the PHENIX program and subsequently adjusted manually in COOT. Finally, the atomic model was validated using MolProbity in PHENIX[36,39]. The real-space correlation coefficients calculated for all amino-acid residues are shown as Supplementary Table 3. To avoid overfitting, we validated the final model following a previous method[40]. Three Fourier shell correlation (FSC) curves—that is, model versus final map, FSC_{work} ($model_{sf}$ versus half1 map) and FSC_{free} ($model_{sf}$ versus half2 map)—were produced. The general agreement of these curves was taken as an indication that the model was not overfit. Structural figures were prepared in Chimera[38] and PyMOL (https://pymol.org/2/).

In vitro binding assay of EMC with the TOM complex

The 3× Flag tag was inserted onto the C terminus of the *Tom22* gene. The endogenous *S. cerevisiae* TOM complex was purified by anti-Flag affinity resin and size-exclusion chromatography using the same protocol as for the EMC. In the binding assay, twice as much of the purified TOM was pre-incubated with EMC for 1 h at 4 °C, and then analysed in a Superose 6 10/300 gel filtration column in buffer containing 0.01% GDN, 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, 1 mM $MgCl_2$ and 1 mM $MnCl_2$. As controls, the purified TOM and EMC proteins were analysed separately using the same conditions.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The cryo-EM 3D map of the *S. cerevisiae* EMC complex has been deposited at the Electron Microscopy Data Bank (EMDB) database with accession code EMD-21587. The corresponding atomic model was deposited at the RCSB Protein Data Bank (PDB) with accession code 6WB9. The TMT mass spectrometry data and the real-space correlation coefficients of all residues with experimental densities data are provided in Supplementary Tables 1–3. The uncropped SDS–PAGE gels used in Figs. 1a, 3g and Extended Data Fig. 6b can be found in Supplementary Fig. 1.

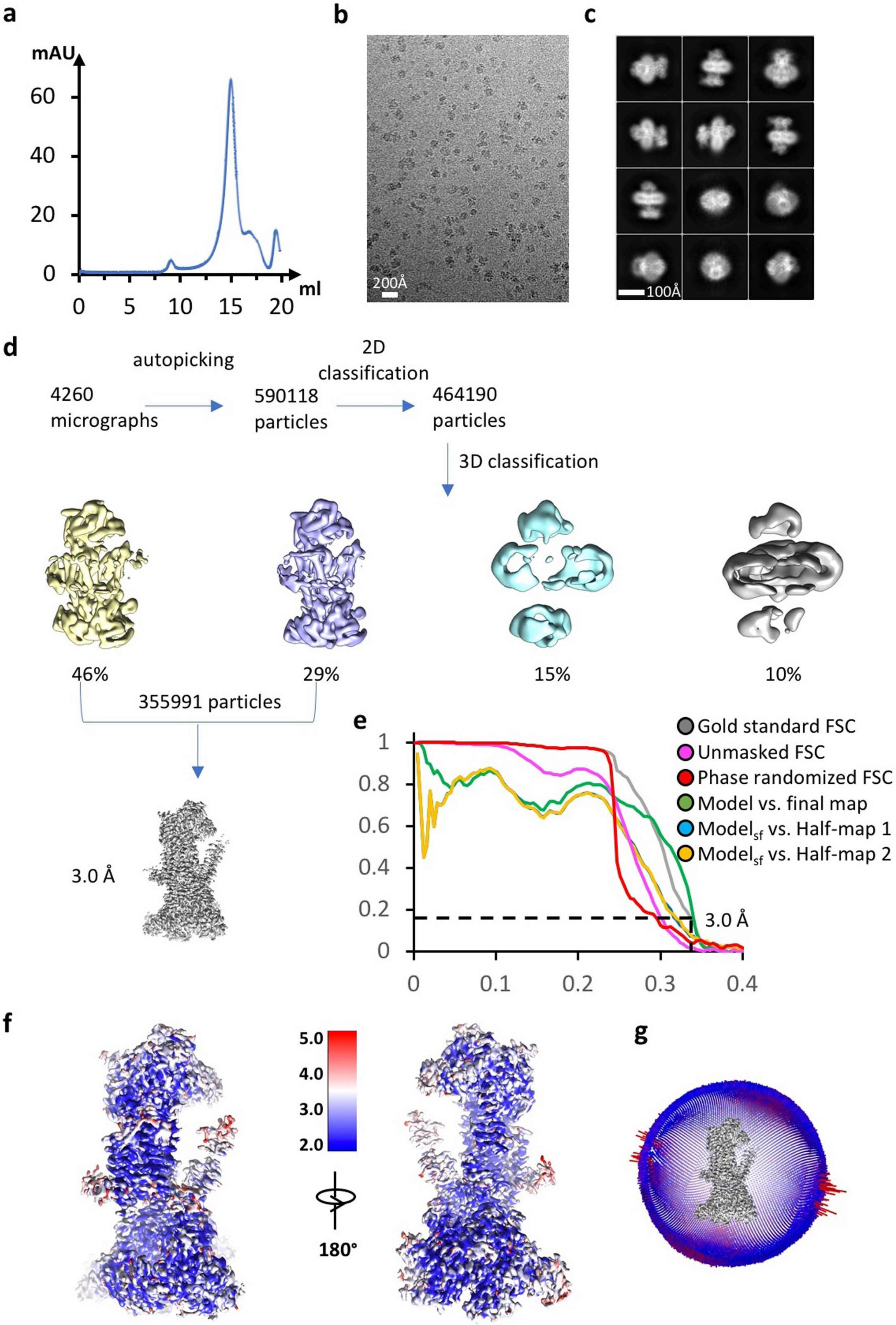
31. Funakoshi, M. & Hochstrasser, M. Small epitope-linker modules for PCR-based C-terminal tagging in *Saccharomyces cerevisiae*. *Yeast* **26**, 185–192 (2009).
32. Kammers, K., Cole, R. N., Tiengwe, C. & Ruczinski, I. Detecting significant changes in protein abundance. *EuPA Open Proteom*. **7**, 11–19 (2015).
33. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
34. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol*. **142**, 334–347 (2003).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
36. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
38. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem*. **25**, 1605–1612 (2004).
39. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
40. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).

Acknowledgements Cryo-EM images were collected in the David Van Andel Advanced Cryo-Electron Microscopy Suite at Van Andel Research Institute. We thank G. Zhao and X. Meng for assistance with data collection and D. Nadziejka for critical reading of the manuscript. This work was supported by the US National Institutes of Health (NIH) (R01 CA231466 to H.L.) and Van Andel Institute (to H.L.).

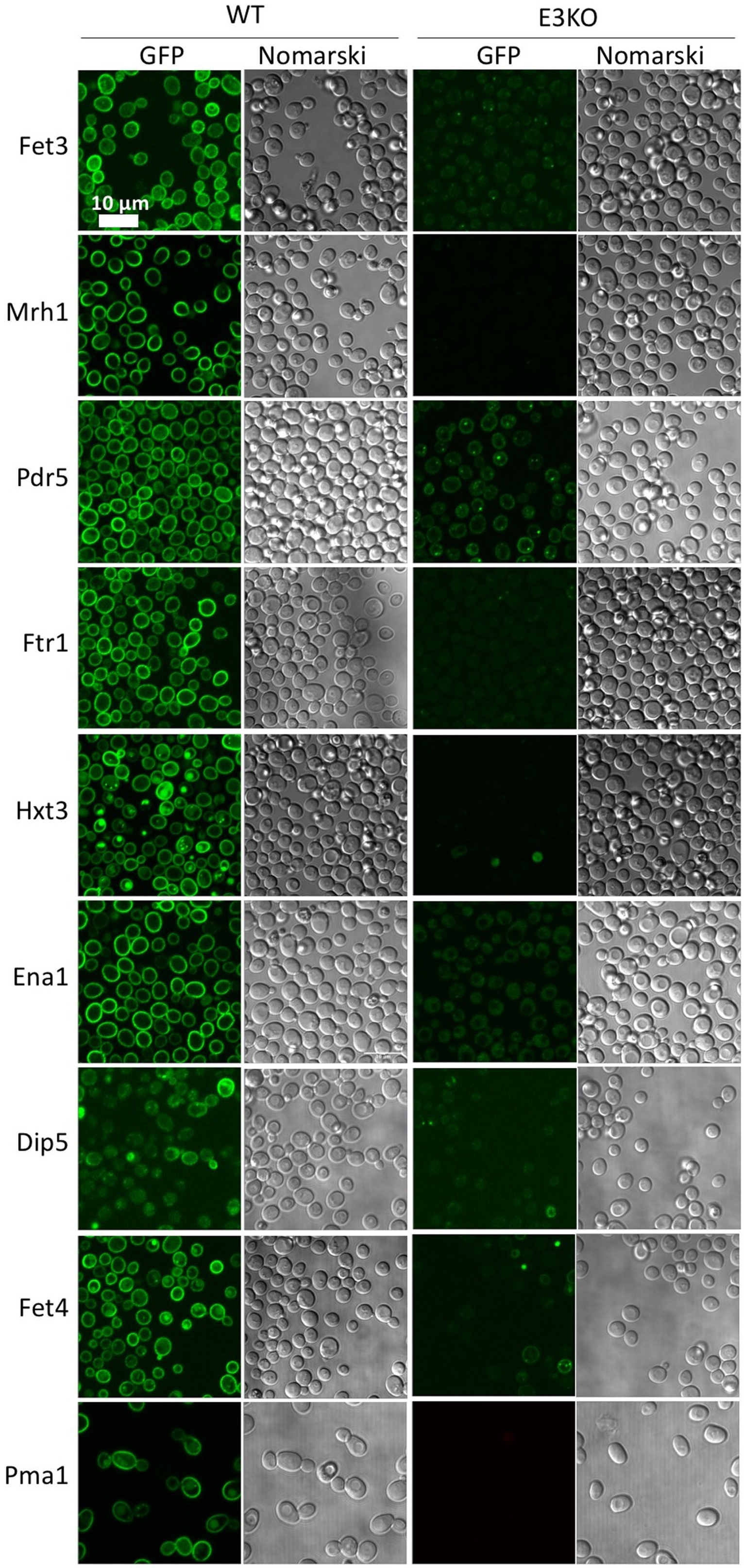
Author contributions L.B. and H.L. conceived and designed experiments. L.B. performed most of the experiments. Q.Y., X.F. and A.K. helped with sample preparation and functional assays. L.B. and H.L. analysed the data and wrote the manuscript.

Competing interests The authors declare no competing interests.

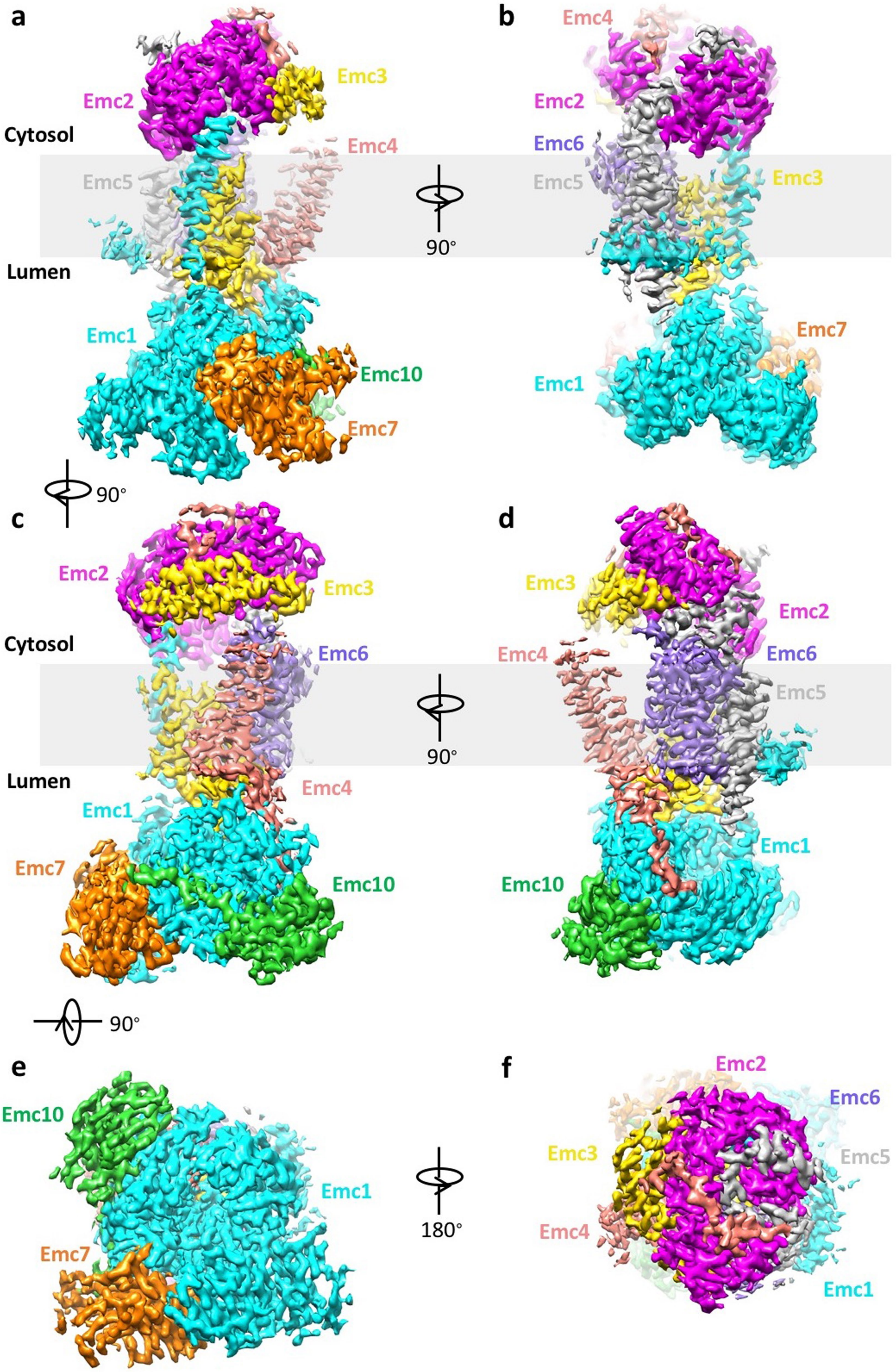
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2389-3.
Correspondence and requests for materials should be addressed to L.B. or H.L.
Peer review information *Nature* thanks Friedrich Förster and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



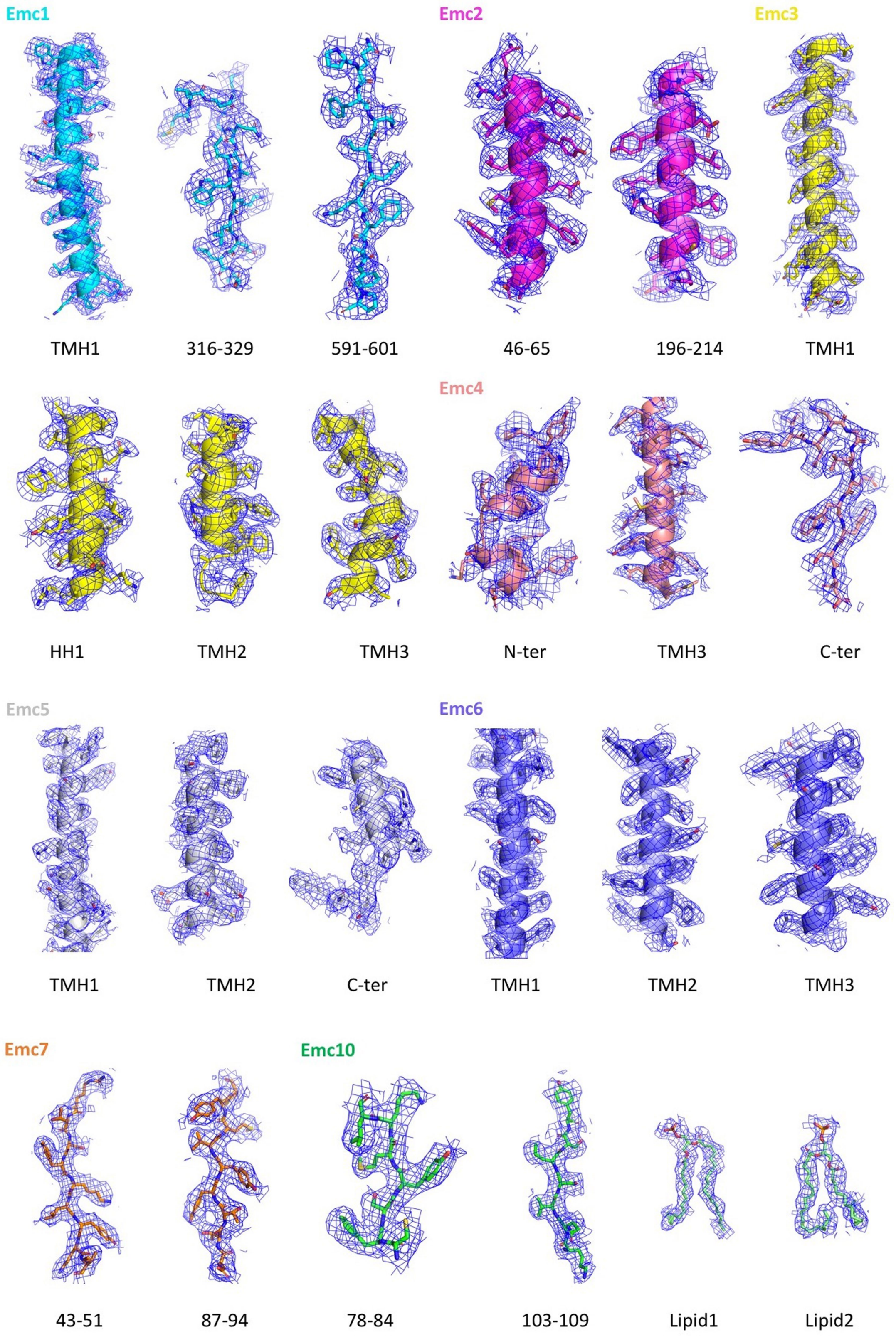
Extended Data Fig. 1 | Data processing and validation of cryo-EM micrographs and 3D reconstruction. a, Gel filtration profile of the EMC complex. This experiment was repeated more than five times with similar results. **b**, **c**, Representative electron micrograph and selected reference-free 2D class averages of the EMC. A total of 4,260 micrographs were recorded with similar quality. **d**, Cryo-EM data-processing procedure. **e**, Gold-standard Fourier shell correlations of two independent half maps with or without mask, and with randomized phases, and the validation correlation curves of the atomic model by comparing the model with the final map or with the two half maps. **f**, Local-resolution map of the 3D map. **g**, Angular distribution of particles used in the final reconstruction of the 3D map.



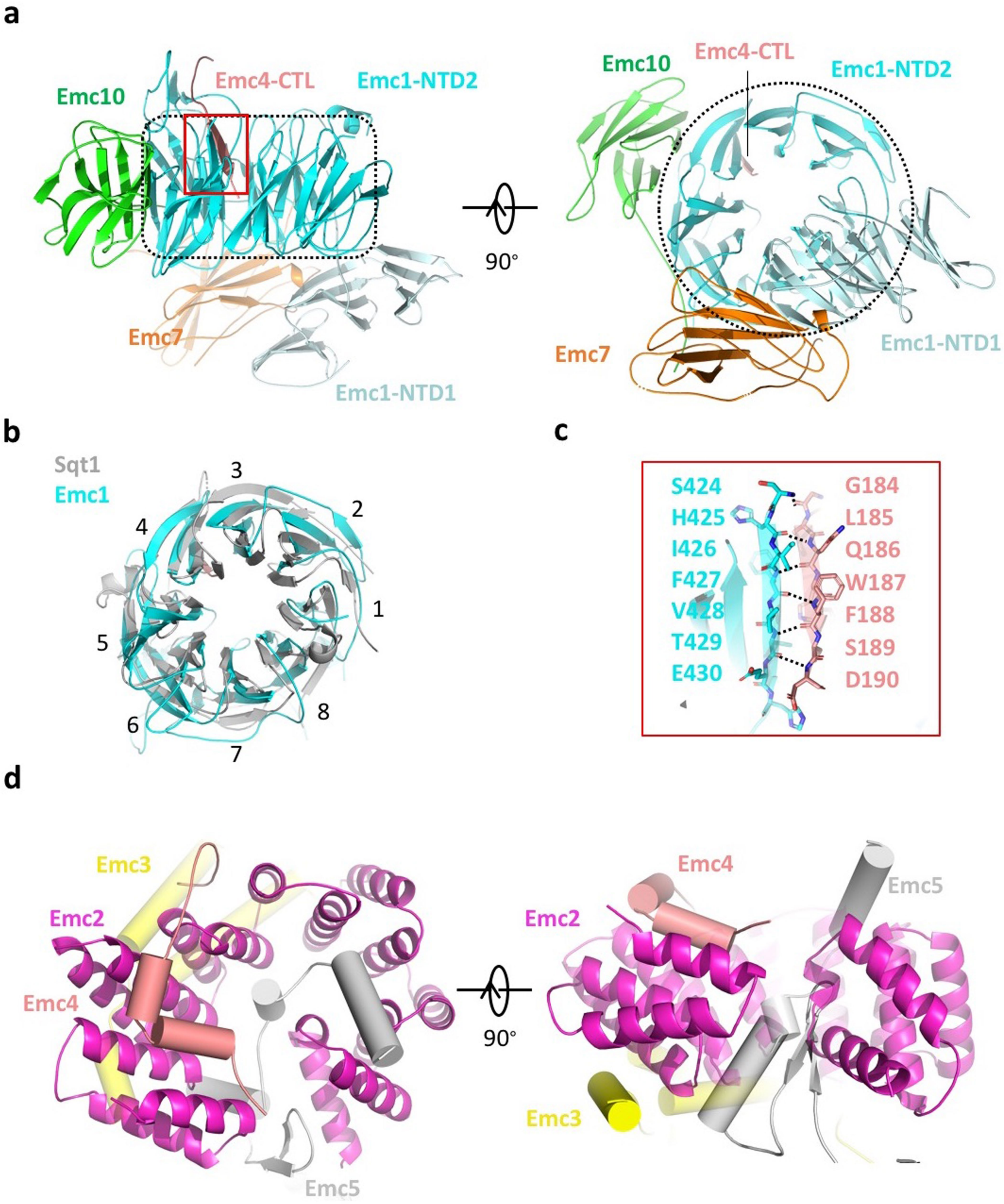
Extended Data Fig. 2 | Protein abundance and localization of nine putative EMC clients in wild-type and Emc3-knockout yeast strains. The eGFP is appended to the C termini of the genes. Scale bar, 10 µm. This experiment was repeated three times with similar results.



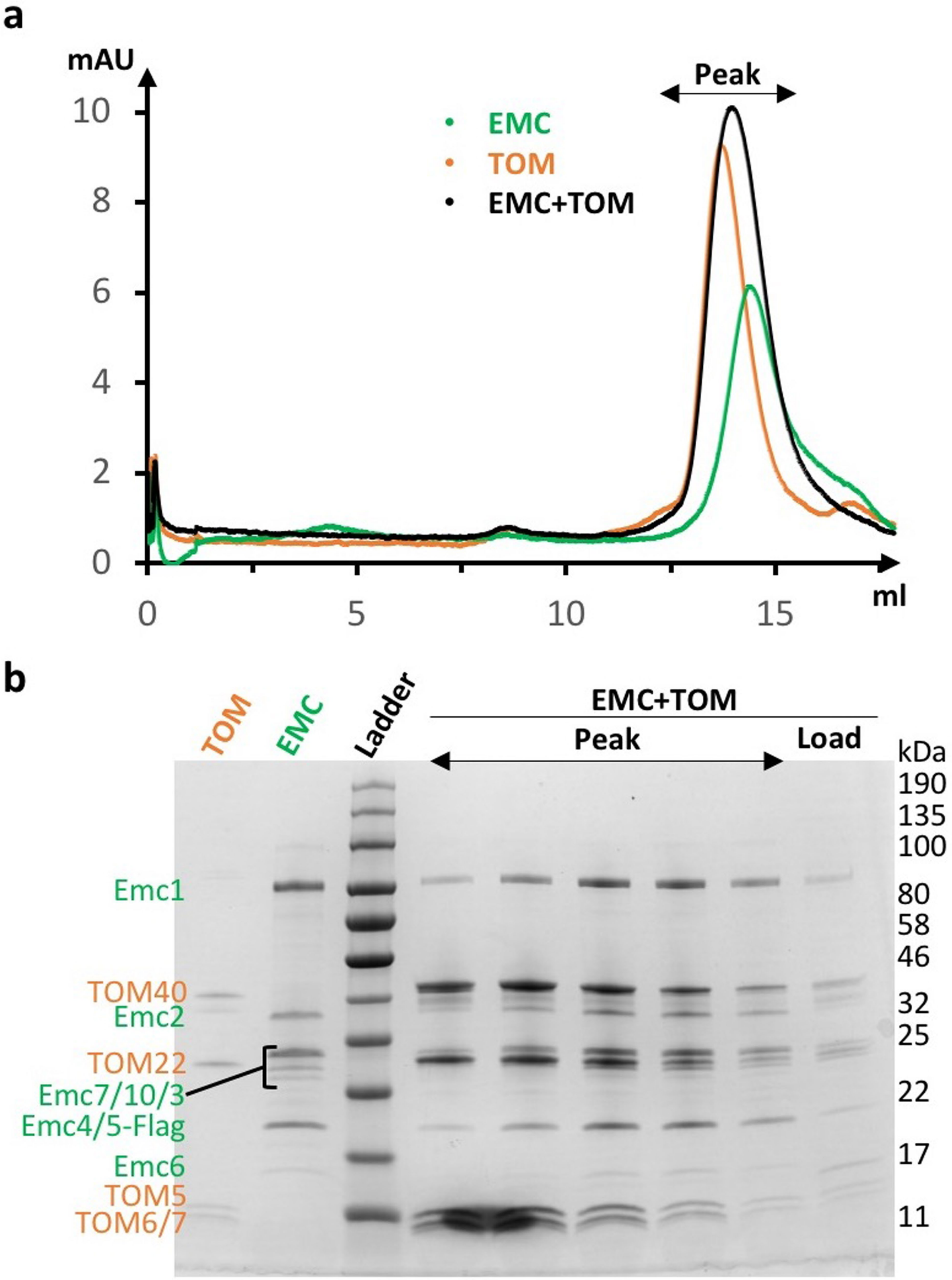
Extended Data Fig. 3 | Cryo-EM 3D density map of the EMC. a–**d**, The surface-rendered map is shown in front view (**a**), left side view (**b**), right side view (**c**), back view (**d**), bottom (lumenal) view (**e**), and top (cytosolic) view (**f**). Maps are coloured by individual subunits.



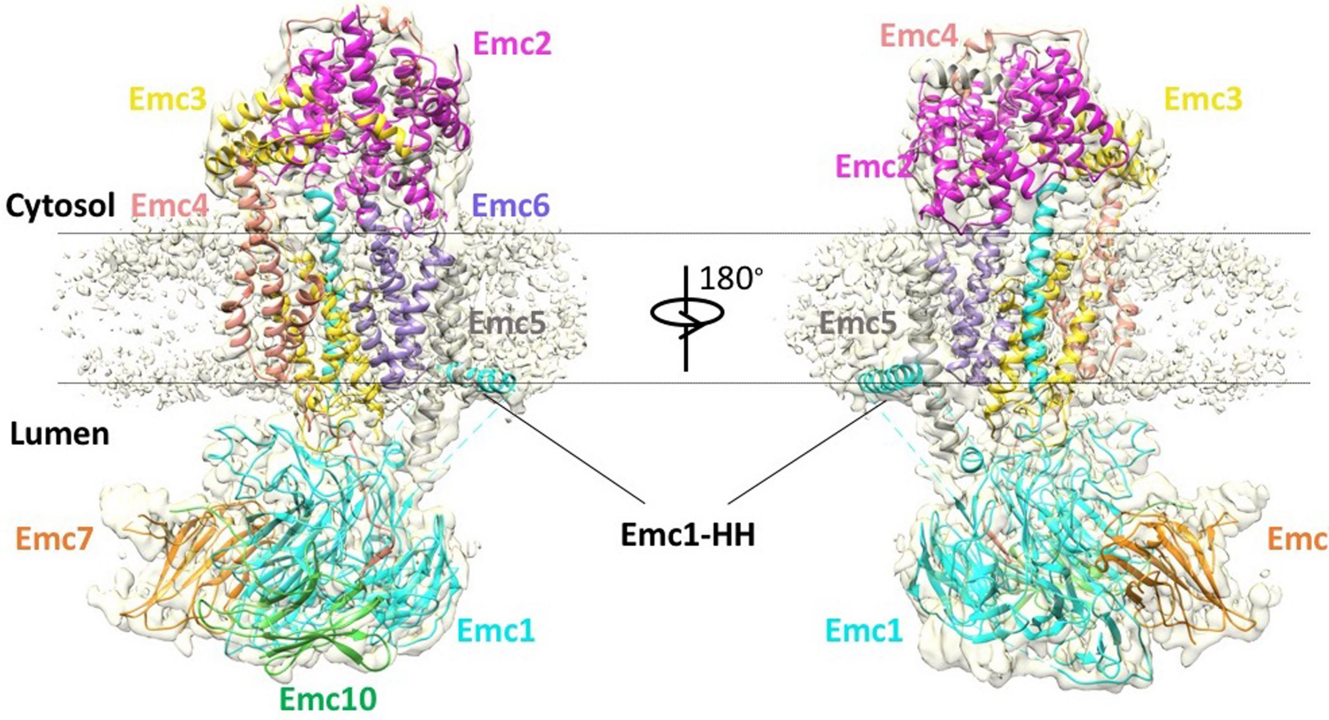
Extended Data Fig. 4 | The fitting of the atomic model and the 3D map in selected regions. 3D density map and atomic model of selected regions in each of the eight EMC subunits, as well as the densities of atomic models of the two phospholipid molecules. C-ter, C-terminal domain; HH, horizontal helix; N-ter, N-terminal domain.



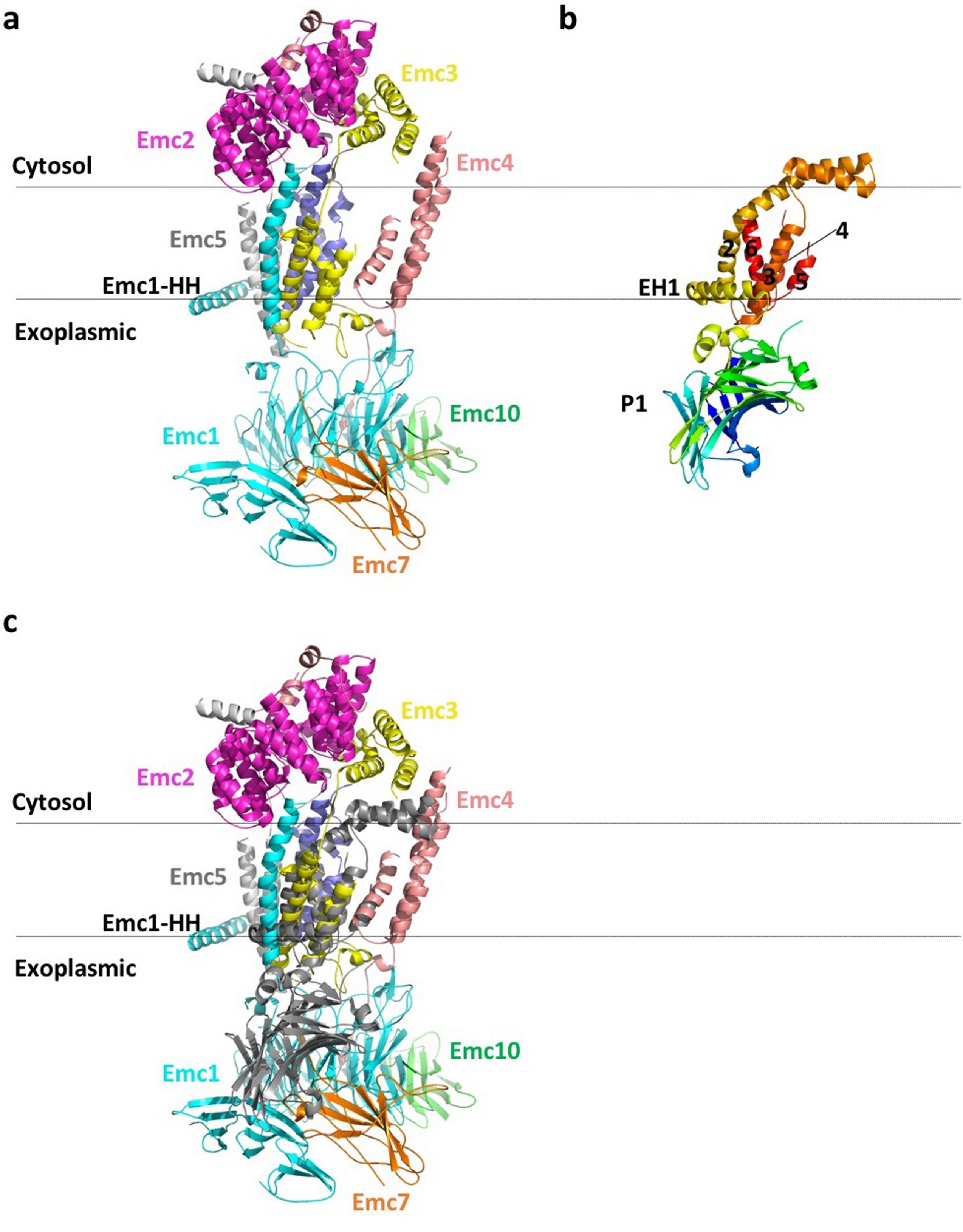
Extended Data Fig. 5 | Structure of the lumenal and cytosolic regions of the yeast EMC. a, Structure of the EMC lumenal region shown in front side and bottom (lumenal) views. The interface area between the C-terminal loop of Emc4 and the NTD2 of Emc1 is outlined by a red rectangle. The dotted black area marks the NTD2 of Emc1, which is an eight-bladed β-propeller. **b**, Superposition of the NTD2 β-propeller of Emc1 with the structure of a fungus chaperone protein Sqt1 (PDB code 4ZN4). **c**, Enlarged view of the red-outlined region in **a**. **d**, Structure of the EMC cytosolic region in top (cytosolic) and front side views. Emc2 as the organizing centre is shown in cartoon, and the cytosolic domains of Emc3, Emc4 and Emc5 are shown as cylinders.



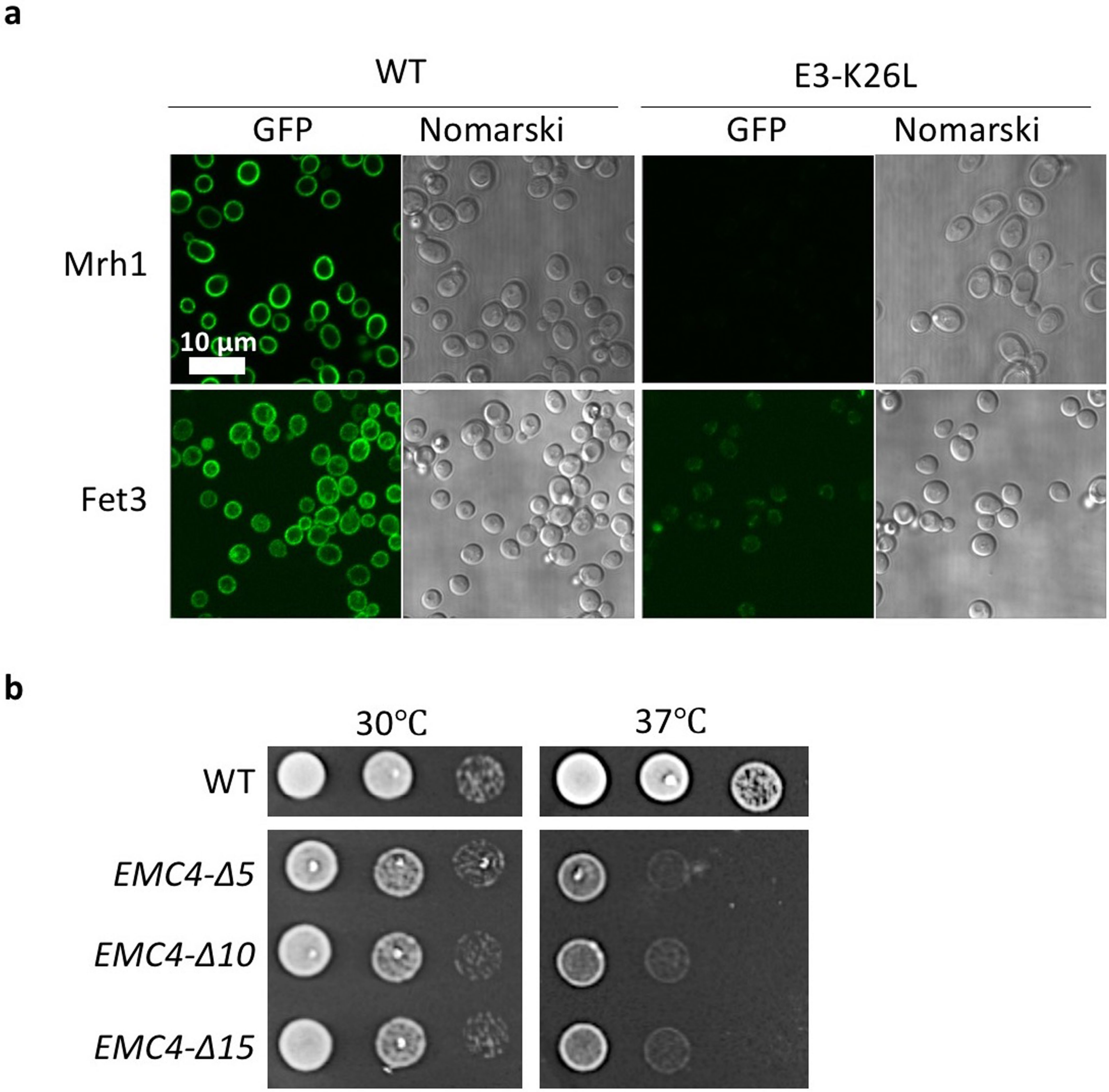
Extended Data Fig. 6 | In vitro binding assays between the purified EMC and the TOM complex. a, Gel filtration profiles of the EMC alone, the TOM complex alone, and the mixture of the EMC–TOM complexes. No peak corresponding to the assembly of the EMC–TOM complex was observed. The experiment was repeated three times yielding similar results. **b**, Peak fractions of the EMC–TOM mixture in **a** were checked by the Coomassie blue-stained SDS–PAGE gel. The band densities suggest that the peak is simply an overlap of the unbound and separate EMC and TOM. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | The 3D EM map of the EMC surface rendered at a low display threshold. The bound lipids/detergents surrounding the transmembrane region of the EMC complex are visible in this low-threshold display. The atomic model in cartoon is superimposed on the 3D map. Note that the horizontal helix of Emc1 is at the ER lumen–membrane boundary.



Extended Data Fig. 8 | Structural comparison between yeast EMC and *E. coli* YidC. **a**, Structure of EMC in cartoon. **b**, Structure of *E. coli* YidC in cartoon (PDB code 3WVF). **c**, Superposed structures of EMC (colour) and YidC (dark grey).



Extended Data Fig. 9 | Comparisons of protein abundance, localization, and growth of the mutant yeast strains with the wild-type cells. a, Protein abundance and localization of two putative EMC clients (Mrh1 and Fet3) in wild-type and EMC3(K26L) mutant yeast strains. The eGFP is appended to the C termini of the genes. **b**, Growth experiments of yeast strains containing Emc4 linker loop truncations. The three truncations were Emc4(Δ56–60), Emc4(Δ51–60), and Emc4(Δ46–60). Experiments in **a** and **b** were repeated three times with similar results.

Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statisticsCryo-EM data collection, refinement, and validation statistics

	S. cerevisiae EMC (EMDB-21587) (PDB 6WB9)
Data collection and processing	**Titan Krios (FEI)**
Magnification	130,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	69
Defocus range (μm)	1.0-2.0
Pixel size (Å)	1.029
Symmetry imposed	C1
Initial particle images (no.)	590,118
Final particle images (no.)	355,991
Map resolution (Å)	3.0
FSC threshold	0.143
Map resolution range (Å)	3.0-250
Refinement	
Model resolution (Å)	3.1
FSC threshold	0.5
Model resolution range (Å)	3.1-250
Map sharpening *B* factor (Å^2)	96.5
Model composition	
Non-hydrogen atoms	14,510
Protein residues	1771
Ligands	8
B factors (Å^2)	
Protein	51.4
Ligand	65.4
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.748
Validation	
MolProbity score	1.99
Clashscore	9.85
Poor rotamers (%)	0.87
Ramachandran plot	
Favored (%)	92.2
Allowed (%)	7.8
Disallowed (%)	0

nature research

Corresponding author(s): Huilin Li

Last updated by author(s): 03/28/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cryo-EM data collection used SerialEM in Titan Krios and EPU as implemented in Arctica by the manufacturer Thermo-Fisher Scientific. Fluorescent microscopy images were collected using program NIS-Elements Software.
Data analysis	RELIONS 3.0, MotionCorr 2.0, CTFFIND 4.1, Chimera, Pymol, Coot, Phenix, and MolProbity, and ImageJ. Mass spectrometry data was analyzed following the protocol and scripts published by Dr. K. Kammers (http://www.biostat.jhsph.edu/~kkammers/software/eupa/R_guide.html).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM 3D map of the S. cerevisiae EMC complex was deposited in the EMDB database with accession code EMD-21587. The corresponding atomic model was deposited in the RCSB PDB with accession code 6WB9.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We collected 4260 raw movie micrographs in Titan Krios. We picked 590,118 raw particles. After 2D classification, a total of 464,190 "good" particles that produced clear 2D class averages were retained in the dataset. After 3D classification, 355,991 raw particles were retained, refinement of which resulted in the final 3D map at 3.0 Å resolution. The sample size was deemed sufficient because the data yielded our targeted resolution of better than 3.5 Å.
Data exclusions	"Bad" raw particles that did not produce 2D class averages or 3D class maps with defined features were excluded after 2D and 3D classifications. This criteria is empirical but is a standard image processing practice in the cryo-EM community.
Replication	Reproducibility resides in the large number of particles used to derive at the final 3D maps or 2D averages. The reliability and the resolution is measured by gold-standard Fourier shell correlation. Replication efforts with multiple refinement runs yielded was successful, yeilding the similar 3D maps.
Randomization	The raw particles were randomly selected by computer program (RELIONS 3.0).
Blinding	The investigators were blinded to the specific data points during data collection and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The yeast strain W303-1a (MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11) was obtained from the Mike O'Donnell lab at Rockefeller University.
Authentication	The strain was not authenticated.
Mycoplasma contamination	The cells were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	N/A.

Structure of the essential inner membrane lipopolysaccharide–PbgA complex

https://doi.org/10.1038/s41586-020-2597-x

Received: 22 March 2018

Accepted: 10 July 2020

Thomas Clairfeuille[1,13], Kerry R. Buchholz[2,13], Qingling Li[3], Erik Verschueren[3], Peter Liu[3], Dewakar Sangaraju[4], Summer Park[5], Cameron L. Noland[1], Kelly M. Storek[2], Nicholas N. Nickerson[2], Lynn Martin[6], Trisha Dela Vega[6], Anh Miu[7], Janina Reeder[8], Maria Ruiz-Gonzalez[9], Danielle Swem[2], Guanghui Han[3], Daniel P. DePonte[10], Mark S. Hunter[10], Cornelius Gati[11,12], Sheerin Shahidi-Latham[4], Min Xu[5], Nicholas Skelton[9], Benjamin D. Sellers[9], Elizabeth Skippington[8], Wendy Sandoval[3], Emily J. Hanan[9]✉, Jian Payandeh[1,2]✉ & Steven T. Rutherford[2]✉

Lipopolysaccharide (LPS) resides in the outer membrane of Gram-negative bacteria where it is responsible for barrier function[1,2]. LPS can cause death as a result of septic shock, and its lipid A core is the target of polymyxin antibiotics[3,4]. Despite the clinical importance of polymyxins and the emergence of multidrug resistant strains[5], our understanding of the bacterial factors that regulate LPS biogenesis is incomplete. Here we characterize the inner membrane protein PbgA and report that its depletion attenuates the virulence of *Escherichia coli* by reducing levels of LPS and outer membrane integrity. In contrast to previous claims that PbgA functions as a cardiolipin transporter[6–9], our structural analyses and physiological studies identify a lipid A-binding motif along the periplasmic leaflet of the inner membrane. Synthetic PbgA-derived peptides selectively bind to LPS in vitro and inhibit the growth of diverse Gram-negative bacteria, including polymyxin-resistant strains. Proteomic, genetic and pharmacological experiments uncover a model in which direct periplasmic sensing of LPS by PbgA coordinates the biosynthesis of lipid A by regulating the stability of LpxC, a key cytoplasmic biosynthetic enzyme[10–12]. In summary, we find that PbgA has an unexpected but essential role in the regulation of LPS biogenesis, presents a new structural basis for the selective recognition of lipids, and provides opportunities for future antibiotic discovery.

In *E. coli*, the outer membrane is an essential structure where LPS resides within the outer leaflet to impart barrier function and immune modulation[1]. Cell division requires the synthesis and transport of millions of new LPS molecules[1,2], which are composed of a lipid A membrane-anchor, core oligosaccharide, and O-antigen. LpxC performs the committed step of lipid A biosynthesis[13], and after the addition of core oligosaccharides, MsbA flips LPS into the periplasmic leaflet of the inner membrane[1,2]. The LptB$_2$FG complex shuttles mature LPS across the periplasm to LptDE, which promotes LPS insertion into the outer membrane[1,2]. The outer membrane contains phospholipids on the inner leaflet and imbalance of the LPS-to-phospholipid ratio compromises outer membrane function and cell viability[2]. Information about LPS physiology within the inner membrane remains limited, and the mechanisms that coordinate its synthesis and transport to the outer membrane are poorly defined.

PbgA is an enigmatic inner membrane protein proposed to assemble as a homotetrameric complex that shuttles cardiolipin across the periplasm to the outer membrane[6–9]. However, recent structural studies did not conclusively establish direct evidence of cardiolipin binding and transport[7,8,14]. We investigated PbgA because it is required for the pathogenesis of *Salmonella*[6], conserved in *Enterobacteriaceae*, and has an unclear role in maintaining the outer membrane[10,11,15–18]. Our PbgA crystal structure revealed an unanticipated lipid A-binding motif that has uncovered a new paradigm in bacterial physiology in which PbgA directly perceives LPS within the inner membrane to control the cellular balance of LPS biosynthesis by regulating levels of LpxC. We also report the characterization of lipid A-targeting synthetic peptides based on PbgA that can inhibit the growth of diverse Gram-negative pathogens.

PbgA is essential for outer membrane integrity

Our uropathogenic *E. coli* (UPEC) *pbgA* deletion (Δ*pgbA*) strain contained a suppressor mutation, which suggests *pbgA* essentiality[6,18]. This strain was cleared from mice, serum sensitive, and sensitized to large antibiotics that normally cannot penetrate the outer membrane

---

[1]Structural Biology, Genentech Inc., South San Francisco, CA, USA. [2]Infectious Diseases, Genentech Inc., South San Francisco, CA, USA. [3]Microchemistry, Proteomics & Lipidomics, Genentech Inc., South San Francisco, CA, USA. [4]Drug Metabolism & Pharmacokinetics, Genentech Inc., South San Francisco, CA, USA. [5]Translational Immunology, Genentech Inc., South San Francisco, CA, USA. [6]BioMolecular Resources, Genentech Inc., South San Francisco, CA, USA. [7]Biochemical & Cellular Pharmacology, Genentech Inc., South San Francisco, CA, USA. [8]Bioinformatics & Computational Biology, Genentech Inc., South San Francisco, CA, USA. [9]Discovery Chemistry Departments, Genentech Inc., South San Francisco, CA, USA. [10]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [11]Bioscience Division, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [12]Stanford University, Department of Structural Biology, Stanford, CA, USA. [13]These authors contributed equally: Thomas Clairfeuille, Kerry R. Buchholz. ✉e-mail: hanan.emily@gene.com; payandeh.jian@gene.com; rutherford.steven@gene.com

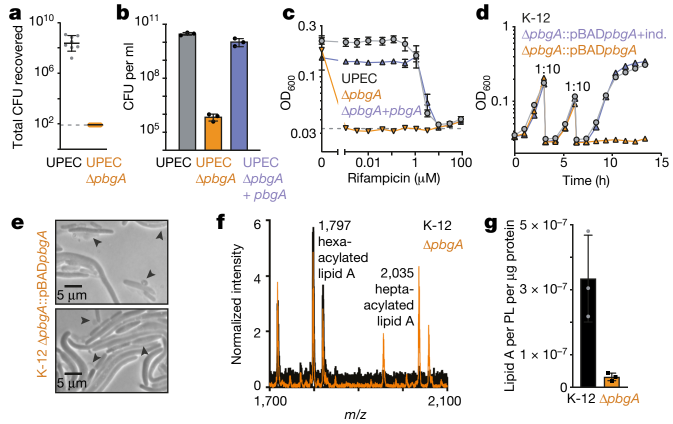


Fig. 1 | PbgA is essential for outer membrane integrity. a, Colony-forming units (CFUs) recovered from the thigh muscle of neutropenic CD1 mice (*n* = 8 per group) 24 h after intramuscular injection. **b**, Strain sensitivity to 50% human serum. **c**, Rifampicin sensitivity of UPEC and UPEC Δ*pbgA* strains diluted into fresh medium containing rifampicin. OD_{600} values were determined at 6 h. **d**, *E. coli* K-12 Δ*pbgA*::pBAD*pbgA* cultures diluted in fresh medium with or without inducer (0.02% arabinose). **e**, *E. coli* K-12 Δ*pbgA*::pBAD*pbgA* grown without arabinose. Images were taken at 4 h and are representative of *n* = 3 experiments. Scale bars, 5 μm. **f**, MALDI–TOF analysis of lipid A extracts from outer membrane vesicles, representative of *n* = 3 experiments. **g**, Quantification of lipid A and phospholipid (PL) by MALDI–TOF analysis of outer membrane vesicles. Data in **a**–**d**, **g** are mean ± s.d. from *n* = 3 independent cultures; line in **a** indicates lower boundary of detection.

(Fig. 1a–c, Extended Data Fig. 1a, Supplementary Table 1). Serum and rifampicin sensitivities were complemented by reintroducing *pbgA* on a plasmid (Fig. 1b, c). In the absence of a suppressor mutation, depletion of PbgA in *E. coli* K-12 resulted in inhibition of growth, rifampicin sensitivity, cells with increased diameter, loss of shape, and membrane bursting (Fig. 1d, e, Extended Data Fig. 1b). Indicating disturbed lipid homeostasis, outer membrane vesicles showed increased hepta-acylated lipid A species and a decrease in the total lipid A:phospholipid ratio relative to wild-type strains[15,19] (Fig. 1f, g). A strain devoid of cardiolipin (Δ*clsABC*) was not sensitized to rifampicin[18,20] (Extended Data Fig. 1c–e). These results establish an essential role for PbgA in pathogenesis, growth and maintaining outer membrane barrier function in *E. coli* in the absence of cardiolipin synthesis.

PbgA is a pseudo-hydrolase

Purified PbgA was monomeric in mild detergent and stabilized by anionic phospholipids, including phospholipid species not naturally abundant in *E. coli* (Extended Data Fig. 2a, b). PbgA crystallized in lipidic cubic phases and the addition of phosphatidylethanolamine allowed high-resolution structure determination (approximately 2 Å), revealing numerous extra densities around the transmembrane domain (TMD) (Extended Data Fig. 2c, d, Supplementary Table 2). PbgA contains five N-terminal transmembrane helices upon which the C-terminal periplasmic domain sits (Fig. 2a). The interfacial domain (IFD) is a compacted three-helix bundle that connects the TMD and periplasmic domain, where substantial interdomain contacts (approximately 2,550 Å^2) suggest the TMD, IFD and periplasmic domain are tightly fused together (Fig. 2a, b, Extended Data Fig. 2e). A distinct crystal form, molecular dynamics studies, and comparison to a recent structure[7] revealed no substantive conformational changes (Extended Data Fig. 3a, b, Supplementary Table 2), indicating that the periplasmic domain remains anchored onto the TMD and protrudes only 60 Å above the inner membrane (Fig. 2a, b). These findings oppose the cardiolipin-transporter model that suggests that the periplasmic domain shuttles across the periplasm[6], which typically measures around 200 Å[21]. Moreover, the IFD is not a simple linker as previously proposed[6], the cardiolipin-binding site hypothesized within the periplasmic domain[8] is distant from the inner membrane and probably cannot permit phospholipid access (Extended Data Fig. 3c), and PbgA is not related to any known transporter (Supplementary Table 3).

PbgA is structurally related to a superfamily of enzymes that modify the cell envelopes of Gram-negative and Gram-positive bacteria (Supplementary Table 3). The periplasmic domain is similar to LtaS, a Mn^{2+}-dependent enzyme that synthesizes an abundant surface polymer in *Staphylococcus aureus*, which lack an outer membrane[22] (Extended Data Fig. 3d). The full-length PbgA structure is most similar to EptA, an inner membrane-anchored, Zn^{2+}-dependent enzyme that transfers a phosphoethanolamine moiety onto lipid A to impart resistance to polymyxin (PMX)[5,23]. Although isolated periplasmic domains and TMDs superimpose well, the compacted α-helical IFD of PbgA exists as an extended linker in EptA, so overall architectures are highly divergent (Extended Data Fig. 3e). Notably, PbgA does not conserve the side chains required to coordinate Zn^{2+} and mutations within its vestigial active site do not affect outer membrane integrity (Fig. 2c, d). Thus, the periplasmic domain appears to be a pseudo-hydrolase, and PbgA has evolved to support an unknown essential function in *E. coli*.

An unanticipated lipid A-binding motif

Strong extra density is observed along the periplasmic membrane leaflet cradled against the IFD of PbgA, but attempts to model or detect cardiolipin failed (Extended Data Figs. 2c, 4a, Supplementary Table 4). Two assays identified the presence of lipid A, and modelling of lipid A rationalized the distinctive bilobal electron density (Extended Data Figs. 2c, 4a–c). Thus, a co-purifying LPS molecule remains bound to PbgA, where the IFD is entirely responsible for coordination using a highly conserved periplasmic lipid A-binding motif (Figs. 2a, b, 3a–d).

PbgA recognizes a minimal feature of lipid A, a single phospho-GlcNAc unit, using eight consecutive residues that precede and form part of the α7-helix, 210-YPMTARRF-217 (Fig. 3b). Specifically, Phe217 anchors the α7-helix within the membrane, and its backbone bonds through water to the *R*-3-hydroxymyristoyl and 1-phospho-GlcNAc of lipid A (Fig. 3b, d). Amides of Arg216 and Arg215 complex with the 1-phospho-GlcNAc, which is further stabilized by the α7-helical dipole (Fig. 3b, d). Notably, the Arg216 side chain is not conserved in all PbgA homologues, and the Arg215 side chain interacts structurally with a conserved acidic residue in the TMD (Extended Data Figs. 2f, 5). Ala214 links to the 210-YPMT-213 segment, allowing the Thr213 backbone to engage the 3-linked *R*-3-hydroxymyristoyl group, and the Thr213 hydroxyl to interact with the 1-hydroxyl and 1-phospho-GlcNAc positions (Fig. 3b). Met212 wedges between the 2- and 3-linked *R*-3-hydroxymyristoyl groups to form hydrophobic contacts (Fig. 3b, d). Pro211 and Tyr210 backbones bond to the 3-linked *R*-3-hydroxymyristoyl substituent, where Pro211 interacts through water (Fig. 3b). Overall, PbgA engages the distinctive chemistry of lipid A using a dense 14-point interaction network primarily through 10 backbone- and water-mediated interactions.

LPS–PbgA interface affects the outer membrane

We introduced point mutations into the PbgA lipid A-binding motif and evaluated outer membrane integrity (Fig. 3c, Extended Data Fig. 6). Charged variants of Met212 imparted sensitivity to rifampicin, in contrast to alanine mutation, which suggests that offending lipid A binding compromised the outer membrane (Fig. 3b–d). Mutation of Thr213 to valine (T213V) did not notably affect rifampicin sensitivity, whereas mutation to aspartic acid (T213D) intended to disrupt the interaction with the 1-phospho-GlcNAc produced extreme sensitization (Fig. 3b, c). Mutation of Arg216 to alanine had no effect, but acidic mutations intended to repulse the 1-phospho-GlcNAc group resulted in rifampicin sensitivities (Fig. 3b, c). The M212A/T213V/R216A triple

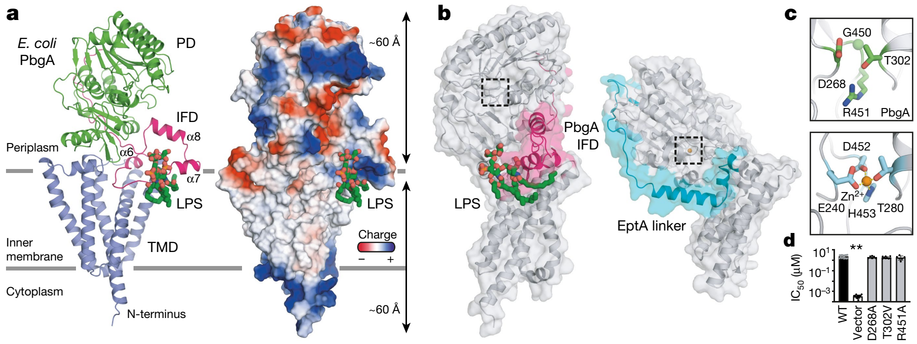


Fig. 2 | PbgA structural features. **a**, PbgA crystal structure in cartoon and electrostatic representation. TMD, IFD and periplasmic domain (PD) are in blue, pink and green, respectively, with LPS as green sticks. **b**, TMD-based alignment with the PbgA–IFD and EptA linker (PDB code 5FGN) in pink and cyan, respectively. Note that the EptA periplasmic domain is oriented approximately 180° relative to PbgA. **c**, Pseudo-hydrolase active site of PbgA (green) and catalytic site in EptA (cyan). **d**, Rifampicin sensitivity of UPEC Δ*pbgA* strains with plasmids expressing wild-type PbgA or mutants. Data are mean ± s.d. from $n = 6$ or more independent experiments per strain. $**P < 0.001$, Bonferroni corrected unpaired two-tailed *t*-test.

mutant produced only a modest phenotype, highlighting the prominent multipoint backbone-mediated coordination scheme observed in PbgA (Fig. 3b–d). Thus, only mutations expected to disrupt lipid A binding along the periplasmic leaflet profoundly affected the outer membrane barrier, which suggests that the LPS–PbgA interface is an essential mediator of outer membrane homeostasis in *E. coli*.

PbgA-derived peptides bind LPS and kill *E. coli*

We postulated that a peptide derived solely from the IFD sequence might bind to LPS in vitro. A synthetic, linear peptide encompassing the lipid A-binding (LAB) motif from PbgA bound to LPS selectively (wild-type LAB (LAB$_{\mathrm{WT}}$); dissociation constant (K_{d}) of approximately 75 μM) over all major *E. coli* phospholipids, whereas peptides expected to destabilize key lipid A-binding determinants (LAB$_{\Delta\alpha7}$ and LAB$_{\mathrm{T213D}}$) showed no binding (Fig. 3e, Extended Data Fig. 7). We predicted that the H221W and D225R mutations might promote membrane partitioning and this LAB$_{\mathrm{WT+}}$ peptide (209-SYPMTARRFLEK<u>W</u>GLL<u>R</u>-225) had improved affinity for LPS (K_{d} value of approximately 55 μM) while maintaining selectivity over phospholipids (Fig. 3e).

Because PMX antibiotics kill Gram-negative bacteria by targeting lipid A[4,5], we tested the LAB peptides for antibacterial activity. The LAB$_{\mathrm{WT}}$ peptide had no effect on *E. coli* growth, potentially owing to its large molecular mass of greater than 2 kDa. For the LAB$_{\mathrm{WT+}}$ peptide, we measured minimal inhibitory concentrations (MICs) of 25–400 μM in chemically or genetically permeabilized cells (Supplementary Tables 5, 6). LAB$_{\Delta\alpha7}$ and LAB$_{\mathrm{T213D}}$ peptides, which were unable to bind LPS, showed no effect on cell growth under matched conditions (Supplementary Table 5). Alanine-scanning and truncation studies ultimately produced a synthetic peptide (LAB$_{\mathrm{v2.0}}$) with an MIC of 200 μM against intact, wild-type *E. coli* K-12 (Supplementary Tables 5, 7, 8).

Optimized LAB achieves broad-spectrum activity

Starting from LAB$_{\mathrm{v2.0}}$, structure-guided design suggested that T213Dap ((*S*)-2,3-diaminopropionic acid) should introduce a salt-bridge to the 1-phospho-GlcNAc, and A214F mutation might improve membrane partitioning and hydrophobic interactions with LPS (Fig. 3b, d). The resulting LAB$_{\mathrm{v2.1}}$ peptide had an MIC of 25 μM against *E. coli* K-12 (Table 1). Inspection of our LPS–PbgA structure and associated data led to three predictions about LAB$_{\mathrm{v2.1}}$ peptide activity (Fig. 3, Extended Data Fig. 4c). First, consistent with conservation of lipid A across Gram-negative bacteria[1], MICs of 12.5–200 μM were obtained against the clinically relevant pathogens *Enterobacter cloacae*,

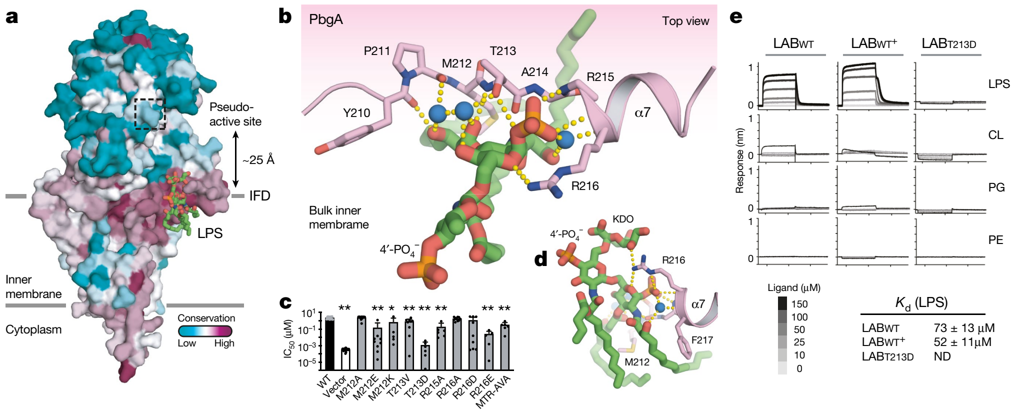


Fig. 3 | The periplasmic lipid A-binding motif of PbgA. **a**, Conservation analysis calculated across 500 PbgA homologues, surface representation. LPS (sticks) and approximate membrane boundaries are indicated. **b**, Close-up view of the lipid A-binding motif with LPS (green stick representation), water molecules (blue spheres) and most bonding interactions (yellow dashes), shown. **c**, Rifampicin sensitivity of UPEC Δ*pbgA* strains with plasmids expressing wild-type PbgA or mutants. Data are mean ± s.d. from $n = 5$ or more independent experiments per strain. $*P < 0.0041$, $**P < 0.001$, Bonferroni corrected unpaired two-tailed *t*-test. MTR-AVA, M212A/T213V/R216A triple mutant. **d**, As in **b**, but a side view. **e**, Synthetic, biotinylated PbgA-derived lipid A-binding (LAB) peptides transferred into different concentrations of detergent solubilized lipids; binding assessed by interferometry measurements. CL, cardiolipin; PE, phosphatidylethanolamine; PG, phosphatidylglycerol. Data are mean and s.d. and representative of $n = 3$ experiments.

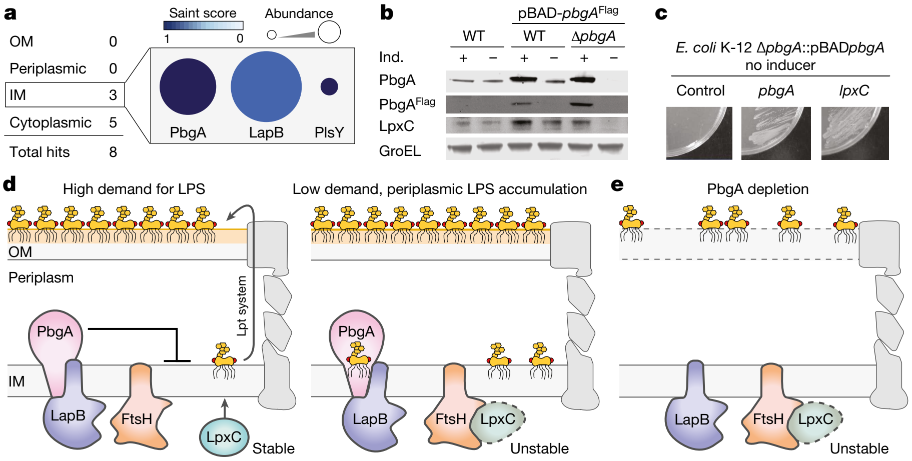


Fig. 4 | PbgA detects periplasmic LPS levels to regulate LpxC stability. **a**, Summary of mass spectrometry analyses following PbgA (endogenous level) immunoprecipitation from *E. coli*. IM, inner membrane; OM, outer membrane. **b**, Western blot of LpxC in the presence or absence of *pbgA*; representative experiment, *n* = 3 or more independent *E. coli* cultures. Ind., inducer. **c**, Growth of conditional PbgA strain with wild-type *pbgA* or *lpxC*; representative plate, *n* = 3 or more independent cultures. **d**, Model of PbgA control of LPS biogenesis and outer membrane integrity. MsbA omitted for clarity and hypothetical cellular states are shown for illustration. When demand for LPS is high, for example, during cell growth (left), the PbgA–LapB complex antagonizes FtsH activity, allowing LpxC to produce LPS precursors. When periplasmic levels of LPS increase, for example, as cells enter stationary phase (right), periplasmic LPS will begin to bind the PbgA–LapB complex, which in turn promotes FtsH degradation of LpxC. **f**, Illustration of the PbgA-depletion phenotype.

Klebsiella pneumoniae, *Acinetobacter baumannii* and *Pseudomonas aeruginosa* (Table 1). Second, consistent with a lipid A-targeting mechanism, growth of the Gram-positive bacterium *S. aureus* that lacks LPS was affected only at very high concentrations (Table 1). Third, when PMX-resistance determinants were introduced into *E. coli*, MICs were unchanged (Table 1, Supplementary Table 9), indicating that LAB peptides and PbgA appear competent to bind unmodified and modified LPS (Fig. 3b, Extended Data Fig. 4c).

The LAB$_{v2.1}$ peptide was bactericidal with time-kill kinetics distinct from PMX antibiotics, potentiated outer membrane-impermeable antibiotics, and synergized with PMX-E (Extended Data Fig. 7b, Supplementary Table 10). Close analogues of LAB$_{v2.1}$ designed to disrupt lipid A interactions had much higher non-specific activity (Extended Data Fig. 7c, d, Supplementary Table 11). Thus, we have discovered a PbgA-inspired class of selective lipid A-binding peptides with activity against Gram-negative pathogens that can overcome modifications that impart PMX resistance.

Table 1 | LAB$_{v2.1}$ peptide exhibits broad-spectrum Gram-negative antibacterial activity

Strain	Phenotype	MICs (μM)[a]
		LAB$_{v2.1}$ YPMXFRRFLEKWGLLR[b]
Escherichia coli ATCC 25922	WT	50
Enterobacter cloacae ATCC 222	WT	12.5
Klebsiella pneumoniae ATCC 43816	WT	100
Acinetobacter baumannii ATCC 19606	WT	12.5
Pseudomonas aeruginosa PA-14	WT	200
Escherichia coli K-12	WT	25
Escherichia coli pmrA^{G53E}	Polymyxin^{R}	12.5
Escherichia coli mcr-1	Polymyxin^{R}	25
*Escherichia coli imp*4213	Permeable	6.25
Staphylococcus aureus USA300	WT	400

[a] MIC is the lowest concentration of compounds that results in complete growth inhibition.
[b] N-terminal acetyl, C-terminal amide, 'X' indicates diaminopropionic acid, a non-natural amino acid.

PbgA controls LPS biosynthesis through LpxC

PbgA immunoprecipitation from *E. coli* identified only two cell envelope hits: the inner membrane proteins PlsY and LapB (Fig. 4a, Extended Data Fig. 8a, Supplementary Table 12). We confirmed PbgA interacts proximally with PlsY and LapB, but not FtsH[24], in intact *E. coli* (Extended Data Fig. 8b). PlsY is involved in phospholipid biosynthesis[25] and LapB has a role in coordinating LPS biogenesis[26–28]. Similar to PbgA (Fig. 1), LapB is essential[26] and its mutation leads to defects in the outer membrane barrier, altered cell morphology and cell bursting[28].

LapB promotes degradation of LpxC, which performs the committed step in lipid A biosynthesis, through modulation of the FtsH protease[24,26]. LpxC was not detected after PbgA depletion, and LpxC levels increased when PbgA was overexpressed (Fig. 4b). Thus, PbgA seems to control LPS levels by functioning as a negative regulator of LapB to ultimately dictate LpxC levels. Accordingly, overexpression of *lpxC* suppressed *pbgA* essentiality, whereas *lapB* overexpression did not (Fig. 4c, Extended Data Fig. 8c).

PbgA is uniquely positioned to detect LPS within the periplasmic leaflet of the *E. coli* inner membrane[10–12] (Fig. 4d). Notably, the PbgA T213D mutant expected to disrupt LPS binding increased LpxC levels and disturbed outer membrane homeostasis (Fig. 3c, Extended Data Fig. 8d). Depletion of periplasmic LPS using an MsbA inhibitor[29] increased levels of LpxC, whereas increasing periplasmic LPS through LptD depletion[30] decreased LpxC levels (Extended Data Fig. 8e, f). We conclude that direct periplasmic sensing of LPS by PbgA controls outer membrane homeostasis through LapB- and FtsH-mediated regulation of LpxC levels.

Discussion

PbgA lacks structural similarity to known transporters or phospholipid-binding proteins. We find that cardiolipin does not co-purify with PbgA, does not bind to the isolated IFD-derived peptide, and is not required to maintain outer membrane integrity in *E. coli*. Our high-resolution crystallographic data permit re-evaluation of a modest PbgA structure[7], which leads to the conclusion that lipid A, not cardiolipin, is bound along the IFD (Extended Data Fig. 9). Moreover, LPS co-purifies with PbgA and binds to the isolated IFD-derived peptide, and lipid A levels are reduced after PbgA depletion, concomitant with a defect in the outer membrane barrier.

PbgA presents a new paradigm in selective lipid recognition as it does not seem to require divalent cations or basic residues to bind lipid A[4,31]. By targeting only a single phospho-GlcNAc unit, PbgA distinguishes itself from known LPS receptors[32], LPS transporters[29,33] and outer membrane proteins[34,35] that exploit the lipid A disaccharide (Extended Data Fig. 10). We leveraged these observations to discover selective LPS-binding peptides that can kill clinically relevant *E. coli*, *E. cloacae*, *K. pneumoniae*, *A. baumannii* and *P. aeruginosa* bacteria in vitro (Table 1), including PMX-resistance strains. Further improvements of $LAB_{v2.1}$ peptide potency, selective outer membrane partitioning, and activity in serum will enable assessment in preclinical infection models.

Exactly how LPS synthesis and transport are coordinated to maintain outer membrane integrity has remained unclear[10–12], but here we reveal the structural basis of an essential LPS–PbgA interaction within the inner membrane. In our model, when cellular demand for LPS is high, LpxC must be stable and active, leading to positive LPS flux (Fig. 4d, left). Under this condition, PbgA exists bound to LapB in an LPS-free state and antagonizes FtsH proteolytic activity. When periplasmic levels of LPS increase, LPS binds to PbgA, altering PbgA–LapB interactions, which promotes activation of FtsH to degrade LpxC (Fig. 4d, right). Overall, LPS levels on the periplasmic leaflet of the inner membrane control the rate of LPS synthesis through direct binding or unbinding to PbgA, functioning as a rheostat to dictate LpxC levels (Fig. 4d).

Our model rationalizes the PbgA depletion phenotype (Fig. 4e) and indicates that disruption of the periplasmic LPS–PbgA interaction may represent a compelling antibacterial strategy. However, key questions persist. LapB remains associated with the PbgA–TMD after deletion of the IFD and periplasmic domain, or when disruptive mutations are introduced into the lipid A-binding motif, which suggests that LapB and PbgA form a constitutive complex (Extended Data Fig. 8g–i). Thus, how LPS binding alters the LapB–PbgA interaction and modulation of FtsH activity remains unknown. A defect in the outer membrane exists in the PbgA–TMD-only strain, indicating altered LPS levels due to an inability to sense LPS, but why this mutant remains viable is not clear[15,18] (Extended Data Fig. 8g–i). A putative phosphatidylethanolamine bound within a conserved cleft on PbgA (Extended Data Fig. 2b) will certainly fuel speculation of a cryptic activity in the TMD[7,14] and other connections to phospholipid biology[6,17,18] (Extended Data Fig. 8j, k). Overall, we have characterized PbgA as a key regulator of LPS biogenesis and outer membrane integrity through the direct detection of LPS on the periplasmic leaflet of the inner membrane, and also present opportunities for future antibiotic discovery.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2597-x.

1. Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem*. **83**, 99–128 (2014).
2. Shrivastava, R. & Chng, S. S. Lipid trafficking across the Gram-negative cell envelope. *J. Biol. Chem*. **294**, 14175–14184 (2019).
3. Parrillo, J. E. Pathogenetic mechanisms of septic shock. *N. Engl. J. Med*. **328**, 1471–1477 (1993).
4. Pristovsek, P. & Kidric, J. Solution structure of polymyxins B and E and effect of binding to lipopolysaccharide: an NMR and molecular modeling study. *J. Med. Chem*. **42**, 4604–4613 (1999).
5. Poirel, L., Jayol, A. & Nordmann, P. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. *Clin. Microbiol. Rev*. **30**, 557–596 (2017).
6. Dalebroux, Z. D. et al. Delivery of cardiolipins to the *Salmonella* outer membrane is necessary for survival within host tissues and virulence. *Cell Host Microbe* **17**, 441–451 (2015).
7. Fan, J., Petersen, E. M., Hinds, T. R., Zheng, N. & Miller, S. I. Structure of an inner membrane protein required for PhoPQ-regulated increases in outer membrane cardiolipin. *MBio* **11**, e03277-19 (2020).
8. Dong, H. et al. Structural insights into cardiolipin transfer from the Inner membrane to the outer membrane by PbgA in Gram-negative bacteria. *Sci. Rep*. **6**, 30815 (2016).
9. Rossi, R. M., Yum, L., Agaisse, H. & Payne, S. M. Cardiolipin synthesis and outer membrane localization are required for *Shigella flexneri* virulence. *MBio* **8**, e01199-17 (2017).
10. Guest, R. L., Samé Guerra, D., Wissler, M., Grimm, J. & Silhavy, T. J. YejM Modulates activity of the YciM/FtsH protease complex to prevent lethal accumulation of lipopolysaccharide. *MBio* **11**, e00598-20 (2020).
11. Fivenson, E. M. & Bernhardt, T. G. An essential membrane protein modulates the proteolysis of LpxC to control lipopolysaccharide synthesis in *Escherichia coli*. *MBio* **11**, e00939-20 (2020).
12. Nguyen, D., Kelly, K., Qiu, N. & Misra, R. YejM controls LpxC levels by regulating protease activity of the FtsH/YciM complex of *Escherichia coli*. *J. Bacteriol.JB*. 00303-20 (2020).
13. Sorensen, P. G. et al. Regulation of UDP-3-O-[*R*-3-hydroxymyristoyl]-*N*-acetylglucosamine deacetylase in *Escherichia coli*. The second enzymatic step of lipid a biosynthesis. *J. Biol. Chem*. **271**, 25898–25905 (1996).
14. Gabale, U., Palomino, P. A. P., Kim, H., Chen, W. & Ressl, S. New functional identity of the essential inner membrane protein YejM: the cardiolipin translocator is also a metalloenzyme. Preprint at https://www.biorxiv.org/content/10.1101/2020.02.13.947838v1 (2020).
15. Hirvas, L., Nurminen, M., Helander, I. M., Vuorio, R. & Vaara, M. The lipid A biosynthesis deficiency of the *Escherichia coli* antibiotic-supersensitive mutant LH530 is suppressed by a novel locus, ORF195. *Microbiology* **143**, 73–81 (1997).
16. Nurminen, M., Hirvas, L. & Vaara, M. The outer membrane of lipid A-deficient *Escherichia coli* mutant LH530 has reduced levels of OmpF and leaks periplasmic enzymes. *Microbiology* **143**, 1533–1537 (1997).
17. Cian, M. B., Giordano, N. P., Masilamani, R., Minor, K. E. & Dalebroux, Z. D. *Salmonella enterica* Serovar Typhimurium uses PbgA/YejM to regulate lipopolysaccharide assembly during bacteremia. *Infect. Immun*. **88**, e00758-19 (2019).
18. De Lay, N. R. & Cronan, J. E. Genetic interaction between the *Escherichia coli* AcpT phosphopantetheinyl transferase and the YejM inner membrane protein. *Genetics* **178**, 1327–1337 (2008).
19. Jia, W. et al. Lipid trafficking controls endotoxin acylation in outer membranes of *Escherichia coli*. *J. Biol. Chem*. **279**, 44966–44975 (2004).
20. Qiu, N. & Misra, R. Overcoming iron deficiency of an *Escherichia coli tonB* mutant by increasing outer membrane permeability. *J. Bacteriol*. **201**, e00340-19 (2019).
21. Fronzes, R. et al. Structure of a type IV secretion system core complex. *Science* **323**, 266–268 (2009).
22. Lu, D. et al. Structure-based mechanism of lipoteichoic acid synthesis by *Staphylococcus aureus* LtaS. *Proc. Natl Acad. Sci. USA* **106**, 1584–1589 (2009).
23. Anandan, A. et al. Structure of a lipid A phosphoethanolamine transferase suggests how conformational changes govern substrate binding. *Proc. Natl Acad. Sci. USA* **114**, 2218–2223 (2017).
24. Ogura, T. et al. Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli*. *Mol. Microbiol*. **31**, 833–844 (1999).
25. Yoshimura, M., Oshima, T. & Ogasawara, N. Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*. *BMC Microbiol*. **7**, 69 (2007).
26. Klein, G., Kobylak, N., Lindner, B., Stupak, A. & Raina, S. Assembly of lipopolysaccharide in *Escherichia coli* requires the essential LapB heat shock protein. *J. Biol. Chem*. **289**, 14829–14853 (2014).
27. Mahalakshmi, S., Sunayana, M. R., SaiSree, L. & Reddy, M. yciM is an essential gene required for regulation of lipopolysaccharide synthesis in Escherichia coli. *Mol. Microbiol*. **91**, 145–157 (2014).
28. Nicolaes, V. et al. Insights into the function of YciM, a heat shock membrane protein required to maintain envelope integrity in *Escherichia coli*. *J. Bacteriol*. **196**, 300–309 (2014).
29. Ho, H. et al. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196–201 (2018).
30. Wu, T. et al. Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 11754–11759 (2006).
31. Lemmon, M. A. Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol*. **9**, 99–111 (2008).
32. Park, B. S. et al. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458**, 1191–1195 (2009).
33. Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the $LptB_2FGC$ complex. *Nature* **567**, 486–490 (2019).
34. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K. & Welte, W. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282**, 2215–2220 (1998).
35. Arunmanee, W. et al. Gram-negative trimeric porins have specific LPS binding sites that are essential for porin biogenesis. *Proc. Natl Acad. Sci. USA* **113**, E5034–E5043 (2016).

Methods

Bacterial strains and plasmids

To generate pBAD-*pbgA*, *pbgA* was amplified from uropathogenic *E. coli* (UPEC CFT073) and cloned into pBAD vector using Gibson assembly according to manufacturer's instructions (New England Biolabs). Mutations in *pbgA* were created using QuikChange II XL site-directed mutagenesis kit (Agilent Technologies) and confirmed by PCR and DNA sequencing.

Mutant strains were created using λ Red recombination[36]. In brief, the kanamycin or gentamicin cassette from pKD4 was amplified with primers containing ~50 bp nucleotide homology extensions to the gene of interest. The linear product was transformed into the appropriate background strain containing pSIM18[37], recovered for 4 h at 37 °C, and selected on medium containing 50 μg ml^{-1} kanamycin or 12.5 μg ml^{-1} chloramphenicol or 10 μg ml^{-1} gentamicin, as appropriate. Mutations were confirmed by PCR and sequencing. Construction of the UPEC-Δ*pbgA* and K-12-Δ*pbgA* strains resulted in single clones and the *pbgA* deletions were confirmed by PCR. Because *pbgA* is reported to be essential[18], we isolated genomic DNA using the Blood and Cell Culture DNA Maxi kit (Qiagen) and sequenced it using the Ilumina HiSeq 2000 platform to identify the suppressor. Paired-end 75 bp reads were aligned to the *E. coli* CFT073 genome using GSNAP version 2013-10-10 with the following parameters: -M 2 -n 10 -B 2 -i 1–pairmax-dna = 1000–terminal-threshold = 1000–gmap-mode = none–clip-overlap. Variant calling was performed using an in-house bioinformatics pipeline using R and Bioconductor packages, GenomicRanges[38], GenomicAlignments[38], VariantTools, and gmapR, with a required base quality score for variant tallying of 30. No single-nucleotide variants or indels were found, but mapping confirmed this strain lacked the *pbgA* gene and identified a large (~569 kb) genomic duplication that straddles the origin (nucleotide positions 1-257,753 and 4,930,864-5,242,376). The mechanism of *pbgA* suppression in this strain has not yet been determined, but *acpT*, a reported multi-copy suppressor of Δ*pbgA*[18], is not duplicated in UPEC Δ*pbgA*.

The conditional *pbgA* strain, Δ*pbgA*::pBAD*pbgA*, was created by inserting pBAD*pbgA* at the *attB* site in BW25113 followed by deletion of the native copy of *pbgA*[36,39]. In brief, *pbgA* was cloned into pBAD28 using standard methods. pBAD*pbgA* was amplified from pBAD28-*pbgA* and sub-cloned into pLDR9. pLDR9-pBAD*pbgA* was digested with NotI, ligated, and transformed into BW25113 expressing pLDR8. PCR and DNA sequencing confirmed insertion of pBAD*pbgA* at the *attB* site. After integration of pBAD*pbgA*, the native copy of *pbgA* was deleted using λ Red recombination as described above.

The triple Δ*clsABC* mutant was constructed by sequentially introducing each individual *cls* deletion from the Keio collection[40] into *E. coli* BW25113 by P1*vir* transduction using standard procedures[41]. Deletions were confirmed by PCR.

pFhuAΔC/Δ4L (pGNE30) was constructed by synthesizing the *fhuA* coding sequence lacking the N-terminal cork domain, Δ1-160, and extracellular loops L3, L4, L5 and L11[42]. *fhuA*ΔcΔ4L was amplified with primers N3P-105 (encoding the *bla* constitutive promoter, ribosome binding site, and AUG start codon from pUC19 (New England BioLabs) and N3P-107, and cloning into pACYC184 with BamHI and HindIII (New England BioLabs). Constitutive expression of *fhuA*ΔC/Δ4L in wild-type *E. coli* results in increased sensitivity to vancomycin and rifamycin.

For complementation and suppression of *E. coli* K-12 Δ*pbgA*::*pBADpbgA*, ASKA (GFP-) plasmids ECK1275 (*lapB*), ECK4026 (*malE*; as control) ECK2182 (*pbgA*), ECK3049 (*plsY*), ECK3459 (*acpT*), ECK2561 (*acpS*), ECK0097 (*lpxC*) were used[43]. Colonies were selected on LB agar plus 25 μg ml^{-1} chloramphenicol and 0.02% arabinose. To test for complementation or suppression, plasmid-containing strains were streaked onto LB agar plates with 25 μg ml^{-1} chloramphenicol but without IPTG as leakiness of the promoter was sufficient to complement (*pbgA)* and higher induction was lethal, or with 20 or 50 μM IPTG (*lpxC*), 50 or 100 μM IPTG (*acpT* and *acpS*) or with all previous listed conditions (*lapB* and *plsY*). For western blot analysis, bacteria scraped from LB agar plates with arabinose were diluted to OD$_{600}$ of 0.025 in LB and arabinose or IPTG conditions as above, grown at 37 °C with aeration, and collected as described below. All strains, plasmids and primers used in this study are listed in the Supplementary Tables 13–15.

Bacterial growth conditions

LB (broth or agar) or Mueller Hinton II cation-adjusted broth (MHB II, BBL 212322) was prepared according to manufacturer's instructions and supplemented with arabinose at 0.02% or at indicated concentrations in figure legends. Bacterial cultures were grown at 37 °C, static, with humidity in 96-well plates for time course and sensitivity assays. To deplete PbgA from Δ*pbgA*::pBAD*pbgA* for western blot analysis, bacteria were grown at 37 °C for around 5 h with dilution to maintain log phase (~8–10 generations), in shaking liquid culture. To deplete PbgA for growth curves, cultures were grown statically at 37 °C and back-diluted 1/10 to maintain logarithmic growth. When appropriate, medium was supplemented with kanamycin (50 μg ml^{-1}), carbenicillin (50 μg ml^{-1}), chloramphenicol (12.5 or 25 μg ml^{-1}), hygromycin (200 μg ml^{-1}), and/or gentamicin (10 μg ml^{-1}). To deplete LptD from *E. coli* K-12 Δ*lptD*::pBAD-*lptD*, bacteria were scraped from LB agar with 0.02% arabinose, diluted into LB broth to an OD$_{600}$ of 0.05, supplemented with 0.02, 0.002 or 0.0002% arabinose, and grown to log-phase at 37 °C with shaking. Bacterial cells were obtained as described below.

Rifampicin-sensitivity assay

For *E. coli* K-12 Δ*pbgA*::pBAD-*pbgA*, bacteria were grown on LB agar plates containing 0.02% arabinose overnight at 37 °C. Cells were scraped from the plate into LB broth, diluted to OD$_{600}$ 0.025, grown to mid-log phase (2.5 h, or approximately 4 generations) at 37 °C, back-diluted in fresh LB broth to OD$_{600}$ 0.025, and grown to mid-log phase again to deplete PbgA accumulated during overnight growth. Rifampicin assay plates were made by serially diluting rifampicin (Sigma) stock (10 mM in DMSO) in LB medium in clear round-bottom 96-well plates (Costar). Bacteria were added to each well to a final OD$_{600}$ 0.01. Plates were incubated at 37 °C statically and OD$_{600}$ read at 4–6 h.

For UPEC strains, each *pbgA*-encoding pBAD28 plasmid was freshly transformed into UPEC-Δ*pbgA* by standard methods and plated onto LB agar plates containing 0.2% glucose and 50 μg ml^{-1} carbenicillin and incubated overnight at 37 °C. Three isolated colonies were picked and re-streaked onto LB agar plus 0.2% glucose and 50 μg ml^{-1} carbenicillin. A single isolated colony from each plate was heavy-streaked onto LB agar containing 50 μg ml^{-1} carbenicillin. Uninduced expression of *pbgA* from the arabinose-inducible promoter was sufficient to complement Δ*pbgA* and PbgA protein levels for each mutant were confirmed by western blot analysis (Extended Data Fig. 6a). Bacteria were scraped from the plate into LB media, diluted, and added to the rifampicin assay plate as described above. Dose–response curves were fit using PRISM software using '[inhibitor] vs response - variable slope' analysis. IC$_{50}$ values from at least four biological replicates were averaged and standard deviation calculated. Values were compared with unpaired two-tailed *t*-test in PRISM and corrected for multiple comparisons (Bonferroni).

Western blot analysis

For UPEC strains, an equivalent of 0.5 OD$_{600}$ bacterial cells were collected by centrifugation and frozen. Pellet was thawed, suspended in PBS and 1× NuPAGE LDS sample buffer (Invitrogen), incubated 20 min, and bath sonicated 10 min in thin-walled sample tubes. Samples were separated on 4–12% NU-PAGE gel (Invitrogen) and transferred to nitrocellulose using the iBLOT2 system (Thermo Fisher Scientific). Nitrocellulose was blocked (PBS with 5% non-fat milk, 0.05% Tween 20) for 1 h and probed for PbgA–Flag overnight at 4 °C with mouse anti-Flag antibody (Cell Signaling Technology) at 1:500–1:1,000 in PBS. A horseradish peroxidase (HRP)-conjugated secondary antibody

(GE Healthcare) at 1:5,000 dilution was incubated with the nitrocellulose for 1 h in 5% non-fat milk in TBS plus 0.05% Tween 20. Between all steps the membrane was washed three times with TBS plus 0.05% Tween 20 and blots were developed using ECL Prime Western Blotting Detection Reagent (Amersham). Blots were stripped with Restore PLUS Western Blot stripping buffer (Thermo Scientific), washed with PBS three times, blocked and probed as described above but with 1:25,000 rabbit anti-GroEL (Enzo) for 1 h.

For all other western blots, an equivalent of OD$_{600}$ of 0.5 bacterial cells from shaking liquid cultures or scraped from LB agar plates were collected by centrifugation and frozen. Pellets were thawed, suspended in 1× LICOR protein sample buffer with 4% β-mercaptoethanol, incubated 10 min at room temperature and then for 5–10 min at 95 °C. Samples were separated on 4–12%, 10%, or 12% NU-PAGE gels (Invitrogen) and transferred to nitrocellulose using the iBLOT2 system (Thermo-Fisher Scientific). Nitrocellulose was blocked in Odyssey PBS blocking buffer (LICOR) for 1–3h at room temperature. Primary antibody incubations were performed at 4 °C overnight at indicated concentrations. Rabbit anti-LpxC (LSBio) was used at 1:5,000–1:10,000 and anti-PbgA monoclonal antibody (7E7, described further below) was used at 1:500, mouse anti-Flag (Cell Signaling Technologies) at 1:500, rabbit anti-GroEL at >1:10,000, human anti-LptD at 1 µg ml^{-1}, all in PBS overnight at 4 °C. After washing membranes three times with TBS plus 0.05% Tween 20, membranes were incubated in Odyssey blocking buffer plus a 1:10,000 dilution of LI-COR goat anti-mouse, anti-rabbit, or anti-human secondary antibodies (IRDye 680RD, IRDye 800CW) and imaged on a LI-COR Odyssey LCx scanner. Antibody information and unprocessed, uncropped western blot gel images are provided in Supplementary Fig. 1.

MIC and time-kill assays

LAB peptides (Smartox Biotechnology, CPC Scientific, ABclonal, standard solid-phase peptide synthesis) at 10 mM in 50 mM Tris, pH 8, and 100 mM NaCl were diluted in MHB II cation adjusted broth (800 µM top concentration) or LB. Where indicated, EDTA was added to a final concentration of 0.5 mM. For modified MIC assays, log phase cultures growing in LB were diluted to OD$_{600}$ of 0.0002 in a final volume of 10 µl in 384-well plates (Corning). Plates were incubated statically at 37 °C and OD$_{600}$ was read after 20 h on EnVision plate reader. For the potentiation MIC assay, log-phase cultures grown in LB were diluted to an OD$_{600}$ of 0.0002 in a final volume of 50 µl in 96-well plate (Corning) with concentration of peptide and antibiotic as indicated in tables. Growth (OD$_{600}$) was measured after static overnight incubation at 37 °C with humidity using a SpectraMax M5 plate reader.

For the time-kill assay, three independent cultures of wild-type *E. coli* (ATCC 25922) were grown to log-phase before being diluted into indicated concentration of peptide relative to the MIC found in Table 1 (that is, 1× MIC = 50 µM) or polymyxin B and incubated at 37 °C, static with humidity. At times indicated, sample was taken, diluted in PBS, and plated on LB agar. CFUs were counted after overnight incubation.

For experiments with MsbA inhibitor[44], *E. coli imp*4213 was grown to an OD$_{600}$ of 0.3, split into three separate cultures (1 µM G913, 4 µM G913, or an equal volume of DMSO), and incubated at 37 °C for 1h. Bacterial cells were collected and processed for western blot analysis with anti-LpxC and anti-GroEL antibodies as described.

Red blood cell lysis assay

Collection of human blood samples from volunteers was through the Genentech Samples for Science Program and carried out under protocols approved by the Western Institutional Review Board (protocol number CEHS-CP 307.2, IRB tracking number 20080040). No personal or medical history was specified, provided or collected for volunteers.

Peptides were diluted in PBS in a 96-well clear round bottom plate at two times the final concentration in 60 µl per well. Whole heparinized human blood was diluted to 4% in PBS and 60 µl added to the diluted peptides such that the final blood concentration was 2%. Plates were incubated at 37 °C, static with humidity for 30 min or overnight then centrifuged at 600*g* for 3 min, 60 µl of supernatant was removed to a clear flat bottom plate and OD405 read on a SpectraMax M5 plate reader (Molecular Devices).

Bacterial two-hybrid

The bacterial two-hybrid assay used the Bacterial Adenylate Cyclase Two-Hybrid (BACTH) System Kit (Euromedex) and is based on published methods[45,46]. Fusions were made using BACTH plasmids encoding T25 or T18 adenylate cyclase domains to the N- or C-terminal where appropriate to ensure domains were present on the cytoplasmic side of the inner membrane. pKT25-*pbgA* was tested against the following baits: pUT18-*lapB*, *plsY* and *ftsH* and pUT18C-*hisM* and *pbgA*. pKT25-*pbgA* truncated, EptA-TM swap, or point mutant variants were tested with pUT18-*lapB*. The T25 plasmid (pKT25-*pbgA*) and a T18 plasmid were co-transformed into an adenylate cyclase-deficient *E. coli* strain (DHM1) and grown for 1–2 days at 30 °C on LB agar plate with 50 µg ml^{-1} kanamycin, 50 µg ml^{-1} carbenicillin, and 40 µg ml^{-1} X-gal. Interacting proteins that re-constituted the CyaA adenylate cyclase active site by bringing T25 and T18 together formed blue colonies while partners that did not interact led to white colonies. At least three single isolated colonies were re-streaked onto fresh agar plates to confirm the phenotype.

Ethics statement

All mice used in the in vivo studies were housed and maintained at Genentech in accordance with American Association of Laboratory Animal Care guidelines. All experimental studies were conducted under protocols approved by the Institutional Animal Care and Use Committee of Genentech Lab Animal Research in an Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC)-accredited facility in accordance with the Guide for the Care and Use of Laboratory Animals and applicable laws and regulations.

In vivo infections

For the in vivo infection model, 7-week-old A/J mice (Jackson Laboratory) were rendered neutropenic by peritoneal injection of two doses of cyclophosphamide (150 mg kg^{-1} on day −4 and 100 mg kg^{-1} on day −1). On day 0, mice were infected by intravenous injection through the tail vein of 1×10^6 CFU mid-exponential-phase bacteria diluted in PBS. At 30 min and 24 h after infection, bacterial burdens in the liver and spleen were determined by serial dilutions of tissue homogenates on LB plates. Samples sizes were not predetermined, data were not blinded and experiments were not randomized.

For the thigh infection model, 6-week-old CD1 mice (Charles River Laboratories) were rendered neutropenic by peritoneal injection of 2 doses of cyclophosphamide (150 mg kg^{-1} on day −5 and 100 mg kg^{-1} on day −2). On day 0, mice were infected by intramuscular injection in the thigh muscle of 2×10^4 CFU mid-exponential-phase bacteria diluted in PBS. At 24 h after infection, bacterial burdens in the thigh muscle were determined by serial dilutions of tissue homogenates on LB plates. Samples sizes were not predetermined and data were not blinded.

Extraction and detection of membrane phospholipids

Membrane phospholipids were extracted from outer membrane vesicles using a modified Bligh–Dyer protocol[47] as follows: outer membrane vesicles were prepared from (1×10^6 cells) and suspended in 0.9 ml water, 2 ml methanol (Thermo Fisher Scientific) and 0.9 ml dichloromethane (Acros Organics) were added and vortexed, and the organic layer was removed. The process was repeated and extracts were combined and dried under steady nitrogen flow. Dried residue was reconstituted in 50 µl of 50:50 dichloromethane:methanol with 10 mM ammonium acetate and subjected to LC–MS/MS analysis. Then, 30 µl of sample was injected on a MetaSil AQ C18 column (150 × 2.0 mm, 3.0 µm, Agilent) on a HPLC system (Shimadzu). The temperatures of the column oven and autosampler were set at 45 °C and 15 °C, respectively. Flow rate was 0.3 ml min^{-1} and the gradient was held at 40% mobile phase A (methanol containing 10 mM ammonium acetate) for the initial 3 min. Mobile phase B (dichloromethane with 10 mM ammonium acetate) was increased to 85% over 9 min, then further increased to 95% in 30 s and maintained at 95% for 3 min before returning to initial conditions for re-equilibration and subsequent injections. The HPLC was coupled to a 6500+ QTRAP mass spectrometer (Sciex) operated under positive ionization mode with the following source settings: turbo-ion-spray source at 350 °C under N_2 nebulization at 20 psi, N_2 heater gas at 10 psi, curtain gas at 30 psi, collision-activated dissociation gas pressure was held at medium, turbo ion-spray voltage at 5,500 V, declustering potential at 20 V, entrance potential at 10 V.

Bacterial membrane lipids phosphatidylethanolamine, phosphatidylcholine and cardiolipin were detected by characteristic head group ions present upon fragmentation in either precursor ion scan mode or neutral loss. For phosphatidylethanolamine and phosphatidylcholine, ions were scanned for neutral loss in positive polarity for losses of 141 Da and 184 Da, respectively. Cardiolipin was detected through precursor ion scan in negative mode with a precursor of 391.5 Da. Collision energies were set to 24 V (phosphatidylethanolamine and phosphatidylcholine), and −65 V for cardiolipin. Other parameters were as follows (flipped for negative polarity): CXP 16, EP 10, IS 4500, CUR 20 at temperature (TEM) of 150 °C.

Time-lapse microscopy

E. coli K-12 Δ*pbgA*::pBAD-*pbgA* with arabinose-inducible *pbgA* grown overnight on LB with 0.02% arabinose was inoculated into LB lacking arabinose and grown for 4.5h to deplete PbgA. Cells were maintained in log phase until spotting onto a glass bottom culture dish (MatTek Corporation) and overlaid with a 1% agarose pad made with LB or MHB media. Imaging was performed on a Nikon Eclipse TE inverted fluorescence microscope with a 100× (NA 1.30) oil-immersion objective (Nikon Instruments). Images were collected every 2 min using an Andor DR electron-multiplying CCD camera (Andor Technology) using NIS-Elements software (Nikon Instruments). Cells were maintained at 37 °C during imaging with a temperature-controlled environmental chamber (World Precision Instruments). A representative image of the morphology changes seen in the time course (time taken indicated in figure legend) and in the >3 biological replicates is shown in the figure.

Recombinant protein expression and purification

Full-length (residues 1–586) of *E. coli* and *S. typhimurium* PbgA followed by a TEV cleavage site, 2×Flag tag and a hexahistidine tag at the C terminus were cloned into a modified pET52b vector. Proteins were expressed in *E. coli* BL21-Gold(DE3) for 48 h in TB autoinduction medium at 17 °C. Fifty grams of cell pellet was resuspended in 250 ml of 50 mM Tris pH 8, 300 mM NaCl, 1 μg ml^{-} benzonase, 1 mM PMSF and Roche protease inhibitor tablets. Cells were lysed by sonication and PbgA were subsequently solubilized by addition of either 1% (w/v) LMNG or 1% (wt/v) dodecyl maltoside (DDM) for 2 h at 4 °C under gentle agitation. Insoluble debris was pelleted by centrifugation at 18,000 rpm for 1 h, and the supernatant containing the solubilized protein was collected for affinity purification by batch-binding to 20 ml of M2-agarose Flag resin (Sigma) for 2 h at 4 °C. Unbound proteins were washed with 10 column volumes of purification buffer (50 mM Tris pH 8, 300 mM NaCl, 0.025% (w/v) LMNG or DDM) and eluted with 5 column volumes of purification buffer supplemented with 150 μg ml^{-1} Flag peptide (Sigma). The eluate was collected and concentrated with 100 kDa MWCO concentrators to 1 mg ml^{-1} before tag removal by TEV cleavage overnight at 4 °C. PbgA was then concentrated to 4 mg ml^{-1}, supplemented with 1 mM $NiCl_2$, and injected onto a Superdex S200 Increase 10/300 column attached to an AKTA system (GE Healthcare) for size-exclusion chromatography into crystallization or SEC-MALS buffer (20 mM sodium citrate pH 5, 200 mM NaCl, 0.025% LMNG or DDM). Elution fractions corresponding to monomeric PbgA in LMNG were pooled and concentrated to 40 mg ml^{-1} for crystallization. For the preparation of *E. coli* MsbA (residues 1–582) and *E. coli* Lnt (residues 1–594), constructs were similarly cloned and proteins were expressed and purified in LMNG using the above protocol. For the purification of LPS-free MsbA (MsbA$_{293}$), *E. coli* MsbA (residues 1–582) was cloned into a pRK vector behind a CMV promoter and transiently transfected into Expi293 cells (human embryonic kidney cells; Thermo Fisher Scientific, A14527) using standard protocols. This cell line was not authenticated, but tested negative for mycoplasma contamination. Following expression in this eukaryotic host, purification of MsbA$_{293}$ was carried out as described above using an endotoxin-free AKTA system.

Crystallization, data collection and structure determination

Crystal screens in LCP were set up using 40 mg ml^{-1} PbgA and a monoolein (Sigma): phosphatidylethanolamine (*E. coli* PE, Avanti Polar Lipids) 99.5:0.5% m/m mixture at 40% hydration. Protein–lipid mixes were prepared at room temperature as previously described[48] and crystals grew in 50 nl drops surrounded by 800 nl reservoir solution. Rounds of optimization in MemMeso HT screens (Molecular dimensions) yielded the best-diffracting PbgA crystals that were obtained in a buffer containing 0.1 M Tris pH 8.0, 0.2 M ammonium sulfate, 40% PEG200 at 4 °C, and grew to their maximum size in approximately 20 days. Crystals were flashed-frozen without further cryoprotection for screening. 180° of X-ray diffraction data were collected from a single crystal at the Stanford Synchrotron Radiation Lightsource beamline SSRL12-2 at 100 K, and integrated and scaled using HKL2000[49]. Diffraction from PbgA crystals was anisotropic; however, treatment through the anisotropy server did not indicate severe anisotropy (https://services.mbi.ucla.edu/anisoscale/)[50], nor lead to noticeable improvement in map quality, so it was not applied. To provide a view of the available diffraction data: quality and completeness across 3 different resolution zones (that is, 2.34–2.3; 2.03–2; and 1.88–1.85) are provided in Supplementary Table 2, where completeness is 62%, $CC_{1/2}$ 0.74, I/σ 1.7 and redundancy 1.9 at 2.0 Å resolution. PbgA crystallized in the *C*2 space group with one monomer in the asymmetric unit. The PbgA structure was determined by molecular replacement using PHASER[51] with the PbgA periplasmic domain search model (PDB: 5I5H). Following rigid-body refinement of the periplasmic domain template, clear electron density was visible for the transmembrane domain. The model was completed manually and rebuilt through iterative refinement and omit maps using COOT[52] and PHENIX[53]. Secondary structure restraints were initially applied during refinement but relaxed, and TLS parameters were also employed at late stages in refinement[54]. LPS was modelled only at very late stages of refinement after all protein, other lipids, and most solvent molecules were accounted for. Because reasonable completeness and data quality were available to 1.85 Å, the structure with ligands were refined against all available data until the last round of refinement, where the resolution was cut back to 2.0 Å (Supplementary Table 2). Conservation analysis was performed with Consurf[55], structural homologues were searched for and identified using the Dali server[56], and all structural figures were generated using PyMOL[57]. Where shown, our density maps were calculated to 2.0 Å with $F_o - F_c$ maps calculated before the inclusion of LPS into the refined model to avoid introducing any bias from this ligand.

Crystallization, data collection and structure determination by serial femtosecond X-ray crystallography

PbgA microcrystals were prepared by optimizing the composition of the precipitant solution the macrocrystals were grown in, eventually yielding 5–10 μm crystals that formed in 0.1 M Tris pH 8.4, 0.2 M ammonium sulfate, 24% PEG200 at 20 °C after 48 h incubation. Crystals were then grown in syringes and prepared for serial femtosecond X-ray crystallography data collection as previously described[58]. LCP-SFX data collection was performed using the CXI instrument at the Linac

Coherent Light Source at SLAC National Accelerator Laboratory. 7.9 MAG was added to the PbgA microcrystals LCP medium at around 30% final concentration, and the mixture was injected at a flow rate of approximately 0.400 µl min^{-1} into a vacuum chamber with a 50 µm diameter nozzle. The X-ray free-electron laser was operated at a repetition rate of 120 Hz at a wavelength of 1.3 Å (9.5 keV), delivering focused X-ray pulses of ~40-fs duration with a FWHM of approximately 1.5 µm in diameter. A total of 556,136 detector frames (corresponding to approx. 80 min of data collection) resulted in an average hit rate of 31%, with a total of 170,725 hits as determined by Cheetah[59]. Diffraction patterns obtained from the hit finding step were fed into the CrystFEL software suite[60] (http://www.desy.de/~twhite/crystfel/) for indexing, integration and final merging from a total of 9,498 crystal diffraction patterns, with an estimated resolution cutoff beyond 4.6 Å, judged by the fall-off of crystallographic figures of merit, such as CC*.

Assignment of the space group $P3_1$ presented an indexing ambiguity, which was resolved using the "ambigator" software package within CrystFEL[61]. After running ambigator on the final data set, the indexing ambiguity did not appear to be perfectly resolved (judging by L-tests, etc.), most probably due to the number of diffraction patterns available for inclusion and the limited resolution of the diffraction patterns. The structure was determined by molecular replacement using PHASER[51] in the $P3_1$ space group with two PbgA monomers in the asymmetric with the PbgA full-length structure as a search model, which had all ligands and solvent removed. The TMD and periplasmic domain domains were refined as independent rigid bodies to allow for conformational flexibility within this different crystal lattice. Conservative refinement procedures were pursued and applying a merohedral twinning with operator [-k,-h,-l] in PHENIX REFINE[53] was ultimately found to return major improvements in map quality and R factors, compared to treatment of the data and refinement in the $P3_121$ space group with one PbgA monomer in the asymmetric unit, which yielded otherwise a similar crystal packing arrangement and overall electron density features. LPS was never refined in the $\text{PbgA}_{\text{XFEL}}$ structure due to the limited data resolution of this crystal form. All structural figures were generated using PyMOL[57] and all density maps were calculated to 4.6 Å, where the $F_o - F_c$ map was calculated before the inclusion of LPS to avoid introducing model bias from this ligand.

Molecular dynamics simulations

An all-atom model of PbgA in a lipid environment was generated from the high-resolution crystal structure using the Protein Preparation function in Maestro[62,63] which adds missing residues, side chains and hydrogens, predicts residue protonation states, and optimizes side chain conformations. LPS atoms without clear density were added using the builder function in Maestro. Two simulations were constructed as follows using the System Builder[63]. A LPS-PbgA simulation contained LPS, protein, lipids, water and ions, whereas a PbgA-only simulation did not contain LPS. In each case, the protein was placed in a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) lipid bilayer. The bilayer was initially aligned manually to the region where the protein surface is most hydrophobic. The system was neutralized with the addition of five chlorine ions in the LPS–PbgA system and 11 chlorine ions in the PbgA-only (LPS removed) system. An orthorhombic box was constructed with a 15 Å buffer around the protein in all dimensions and the regions of the box not occupied by protein or lipid were filled with TIP3P waters. The resulting systems were then equilibrated using the relax_membrane.py[64] and Desmond multisim[65–67] programs.

Following equilibration, two production NPT simulations (LPS–PbgA and PbgA only) were run for 500 ns using Desmond, with a temperature of 300 K, pressure of 1.01325 bar and a 2 ps time-step. To assess whether the simulations had reached equilibrium, two new simulations were run with LPS swapped. Specifically, a second LPS–PbgA system was created using PbgA from the last frame of the 'PbgA only' simulation to which LPS was added. This second 'LPS–PbgA' system was re-equilibrated as described above and then run for an additional 500 ns of production simulation. Similarly, a second PbgA-only simulation was built using PbgA from the last frame of the first LPS–PbgA simulation, this time with LPS removed. The new PbgA-only system was re-equilibrated as described above and then run for 500 ns of production simulation. Protein movement was assessed by calculating the root mean squared deviation (r.m.s.d.) versus time using the Event Analysis tool in Maestro. For each of the four production simulations, the PbgA conformation from each time-step was first aligned to the crystal structure using Cα atoms, then the r.m.s.d. was calculated over all Cα atoms.

Multi-angle laser light scattering

Samples (100 µl) of purified PbgA proteins were injected onto a Waters XBridge BEH 200 column with a flow rate of 0.05 ml min^{-1}. The chromatography system was coupled to a three-angle light scattering detector (mini-DAWN TRISTAR) and a refractive index detector (Optilab DSP, Wyatt Technology). Data analysis was carried out using the ASTRA software. The experimental molar masses of *E. coli* and *S. typhimurium* PbgA (67.6 and 70.9 kDa respectively) were calculated with the protein conjugate analysis tool by subtracting the absorption and scattering contribution of dodecyl maltoside (dn/dc = 0.1435).

Differential scanning fluorimetry

Melting experiments were conducted on a Prometheus NT48 (NanoTemper technologies) by measuring the tryptophan fluorescence 330/350 nm ratio of protein samples concentrated at 0.5 mg ml^{-1} in a standard capillary. Standard deviations were calculated from three independent experiments performed with the same protein sample. Lipids (Avanti Polar Lipids) were mixed with purified PbgA protein at a final concentration of 0.1 mg ml^{-1} and incubated for 30 min at 4 °C before measurement.

Biolayer interferometry

Phospholipid (Avanti Polar Lipids) and Kdo_2-lipid A (US Biological Life Sciences) stock solutions were prepared by resuspension into 25 mM Tris pH 8, 100 mM NaCl, 0.05% LMNG buffer and solubilized overnight at 4 °C. Lipid stocks were diluted before experiments into 25 mM Tris pH 8, 100 mM NaCl, 0.5 mg ml^{-1} BSA, 0.05% LMNG. All assays were performed at 25 °C in 25 mM Tris pH 8, 100 mM NaCl, 0.5 mg ml^{-1} BSA, 0.05% LMNG. Biotinylated-LAB peptides were loaded onto SA biosensors to a response of approximately 0.5 nm. Binding to phospholipids and Kdo_2-lipid A was measured at concentrations of 150, 100, 50, 25 and 10 µM with 300 s association and dissociation steps. Assays were performed in triplicate on an Octet Red384 (ForteBio) and buffer and lipid signals were subtracted by using a biotin-blocked reference streptavidin (SA) biosensor. Dissociation constants for LAB_{WT} and $\text{LAB}_{\text{WT+}}$ interactions with Kdo_2-lipidA were estimated by plotting response values at equilibrium as a function of concentration and fit to a global specific binding with Hill slope model in Prism (Graphpad Software).

Quantification of co-purifying LPS

The LPS content of purified PbgA, MsbA and Lnt proteins (25 ng ml^{-1}) was measured using a Limulus amebocyte lysate (LAL) chromogenic endotoxin quantification assay, according to the manufacturer's instructions (Pierce). A standard curve was generated using LPS from *E. coli* strain O111:B4, as directed by the manufacturer. All proteins were purified in LMNG detergent using matched conditions as described above. One endotoxin unit (EU) was assumed to equal 0.1 ng of LPS.

Extraction of Kdo_2-lipid A and detection by mass spectrometry

The extraction and detection of Kdo_2-lipid A was performed as previously described with minor modifications[68]. Four millilitres of hydrolysis buffer (50 mM sodium acetate hydrolysis buffer pH 4.5) was added to 50 µl of 40 mg ml^{-1} purified PbgA protein in a glass tube. The protein suspension was sonicated for 5 min and then left in a boiling water bath for 30 min to cleave the O-antigen from Kdo_2-lipid A. After cooling to room temperature, lipid A was extracted by addition of 4.5 ml of chloroform (Arcos Organics) and 4.5 ml of methanol (Fisher Scientific). The solvents were vortexed thoroughly and separated by centrifugation at 1,000*g* for 10 min. The bottom organic layer was transferred to a new glass tube. Another 4.5 ml of chloroform was added to the remaining upper phase for the second extraction. After vortex and centrifugation, the bottom layers were combined and dried under a steady stream of N_2 gas. The resulting pellet was then dissolved into 10 µl of methanol–chloroform (1:4, v/v) for MALDI–TOF analysis. The 4800 plus MALDI–TOF/TOF Analyzer (AB Sciex) was equipped with a Nd:YAG laser using a 200 Hz firing rate. The matrix used was a saturated solution of 6-aza-2-thiothymine (Sigma-Aldrich) in 100% methanol. Samples were prepared by depositing 0.5 µl of matrix followed by 0.5 µl of the sample solution on the sample plate. After drying at room temperature, the spectra were acquired in the negative ion reflector mode.

Untargeted metabolomics

PbgA and MsbA samples were diluted to 100 µl reconstitution solvent to a concentration of 0.6 mg ml^{-1} (2:1:1 LC–MS grade water: methanol: acetonitrile) followed by ultra-sonication for 8 min in a room temperature water bath. Five microliters of each sample supernatant was injected for LC–MS analysis. Shimadzu series ultra-high performance liquid chromatography (UHPLC) system (Shimadzu) consisting of LC pumps (Model LC-30AD) with online degasser was used to deliver the mobile phases 5 mM ammonium acetate with 0.1% (v/v) formic acid in water (A) and 1 mM ammonium acetate with 0.1% (v/v) formic acid in acetonitrile:isopropyl alcohol (5:2, v/v) (B) at a flow rate of 0.3 ml min^{-1}. Samples (5 µl) were injected through an autosampler (ModelSIL30ACMP) with temperature control at 15 °C. Kinetex Evo C18 (100 × 2.1 mm 1.7µm; Phenomenex) reverse-phase column was used for liquid chromatography separation. Gradient liquid chromatography flow started with 5% B with a linear increase to 95% B in 30 min, followed by a 95% B hold for 5 min before returning to 5% B for column re-equilibration. The column oven (Model CTO30A) temperature was maintained at 40 °C.

Mass spectrometry analysis was performed on Orbitrap-Q Exactive HF-X instrument (Thermo Fisher Scientific) using Top 10 data-dependent MS^2 analysis based on intensity in both positive and negative modes (separate injections) with background ion exclusion lists. Ion exclusion list for positive and negative modes were created separately using buffer blank sample for dynamic software dependent exclusion of high intensity background ions (top 20 high intensity ions in the first half of LC runtime (0.5–20 min) and another top 20 in the second half (19.5–38 min). Data-dependent scan (dd-MS^2) settings for both positive and negative modes included a full MS scan from mass to charge ratio (*m/z*) of 113.5 to 1,700 at a resolution of 120,000 (full-width at half-maximum, FWHM), automatic gain control (AGC) target value of $1e^6$, maximum injection time (IT) of 200 ms and profile mode data acquisition. MS^2 settings included on the fly, top 10 high-intensity ions MS^2 fragmentation with a scan range of *m/z* 200-2000, resolution of 7,500 (FWHM), AGC target of 5×10^4, maximum IT of 10ms, isolation window of *m/z* 1.5 and profile mode data acquisition. Data-dependent settings included minimum AGC target of 5×10^2, intensity threshold of 5×10^4 with no multiple charge states and dynamic exclusion of 5 s. MS source parameters included Heated Electrospray Ionization (HESI) probe with spray voltage of 3.5 kV (positive mode) or 2.5 kV (negative mode), sheath gas flow rate: 49, auxiliary gas flow rate: 12, capillary temperature: 259 °C and funnel RF level at 80.0.

Data analysis to detect and identify unknown compounds with highest fold difference between PbgA (sample) and MsbA (control) was carried out in Compound Discoverer 2.1.0.401 metabolomics software (Thermo Fisher Scientific) using default workflow of 'Untargeted metabolomics with statistics and detect unknowns with mapped pathways (BioCyc beta) and ID using online databases (chem spider, mzcloud and KEGG)'. Data analysis also included protein purification buffer selected as blank in the analysis. Data processing workflow included default parameters for nodes such as input files, selecting spectra, aligning retention times, detecting and grouping unknown compounds, filling gaps, predicting compositions, searching mzcloud and chem spider databases, mapping to BioCyc (beta) and KEGG pathways, normalizing peak areas and marking background compounds. Sample to control MS peak area ratios and $\log_2$-transformed fold changes were calculated in the data analysis and top identified compounds or molecular formula hits with peak area ratios higher than 5 ($\log_2$-transformed fold change >2.6) are reported in Supplementary Table 4.

Generation of monoclonal antibodies against PbgA

Purified *E. coli* PbgA protein was reconstituted into liposomes for immunization by mixing it 1:1,500 molar ratio with *E. coli* polar lipid extract (Avanti Polar Lipids) overnight at 4 °C in the presence of biobeads for detergent removal. Large multilamellar vesicles were harvested by ultracentrifugation, resuspended in TBS and extruded through a 0.45-µm filter at room temperature. Mouse immunization and hybridoma generation were performed using standard protocols. Culture supernatants were assessed for high-affinity monoclonal antibodies using Octet (Fortebio) with anti-mouse IgG Fc capture biosensors for binding to purified PbgA proteins. Three clones were selected, scaled, and purified by standard methods for the co-immunoprecipitation experiments.

Co-immunoprecipitation and LC–MS/MS

Two milligrams of each antibody was applied to 100 µl MabSelect SuRe protein A resin (GE Healthcare) for 15 min in co-immunoprecipitation (co-IP) buffer (25 mM Tris pH 7.5, 150 mM NaCl, 0.025% LMNG). Unbound antibody was washed twice with 500 µl of co-IP buffer, and the beads were mixed with 50 ml of supernatant containing the matching overexpressed bait protein (according to conditions described in above method section) and incubated under gentle agitation for 2 h at 4 °C. Beads were collected by centrifugation at 2,000*g* for 4 min, then washed twice with 100 µl co-IP buffer, twice with 100 µl co-IP buffer supplemented with 350 mM NaCl, and once again with 100 µl co-IP buffer. Antibody–bait–prey complexes were eluted three times with 100 µl elution buffer (0.1 M glycine pH 3.5, 150 mM NaCl, 0.025% LMNG), and separated from beads and collected by centrifugation off a 0.2-µm filter into collecting tube preloaded with 100 µl 100 mM Tris pH 8 for quick neutralization of the acidic pH. Eluted proteins were separated by SDS–PAGE in Tris-glycine on a 4–20% polyacrylamide gel. Twenty bands per gel lane were excised, washed in 25 mM ammonium bicarbonate (Burdick and Jackson, 100 µl, 20 min), destained with 50% acetonitrile in water (100 µl, 20 min) and reduced with 10 mM dithiothreitol at 60 °C followed by alkylation with 50 mM iodoacetamide at room temperature. Proteins were digested with 0.2 µg trypsin (Promega) in ammonium bicarbonate pH 8 at 37 °C for 4 h. Digestion was quenched with formic acid and the supernatants were analysed directly without further processing by nano LC–MS/MS with a Waters NanoAcquity HPLC system (Waters Corp.) interfaced to a ThermoFisher Fusion Lumos. Peptides were loaded on a trapping column and eluted over a 75 µm analytical column at 350 nl min^{-1} (both columns were packed with Luna C18 resin from Phenomenex). A 30 min gradient was used (5 h total LC–MS/MS time per sample). The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 60,000 FWHM resolution and 15,000 FWHM resolution respectively. The instrument was run with a 3 s cycle for MS and MS/MS.

Proteomics analysis

Tandem mass spectrometric data were analysed using the Mascot search algorithm (Matrix Sciences) against a concatenated target-decoy database comprised of the UniProt *E. coli* K-12 protein sequences

(Taxonomy 83333, downloaded 1 July 2017), known contaminants and the reversed versions of each sequence. Peptide assignments were first filtered to a 1% FDR at the peptide level and subsequently to a 2% FDR at the protein level. Peptide spectral matches (PSMs) per protein were summed per sample across all fractions from the GelC-MS experiment. The Statistical Analysis of INTeractome (SAINT) algorithm (SAINTExpress-spc v.3.6.1)[69] was run with default settings comparing the sum of PSMs for all identified proteins enriched with each antibody separately (target) to the combined pool of control purifications (Supplementary Table 12). Interactions with a SAINT score >0.8 and Bayesian FDR < 0.05 were marked as significant.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Structural data are deposited in the Protein Data Bank (PDB) under accession number 6XLP. All mass spectrometry RAW files were uploaded to the MassIVE data repository, accessible by the identifier MSV000083754, and can be downloaded from ftp://MSV000083754@massive.ucsd.edu. DNA sequencing data were deposited at NCBI under BioProject PRJNA541088, BioSample SAMN11572257, experiment SRX5788703, run SRR9010525. The *E. coli* CFT073 reference genome was deposited at NCBI under BioProject PRJNA624646, BioSample SAMN14575425, accession CP051263. Source data are provided with this paper.

36. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
37. Chan, W. et al. A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic Acids Res*. **35**, e64 (2007).
38. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Comput. Biol*. **9**, e1003118 (2013).
39. Diederich, L., Rasmussen, L. J. & Messer, W. New cloning vectors for integration in the lambda attachment site attB of the *Escherichia coli* chromosome. *Plasmid* **28**, 14–24 (1992).
40. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol*. **2**, 2006.0008 (2006).
41. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, 1972).
42. Mohammad, M. M., Howard, K. R. & Movileanu, L. Redesign of a plugged beta-barrel membrane protein. *J. Biol. Chem*. **286**, 8000–8013 (2011).
43. Kitagawa, M. et al. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res*. **12**, 291–299 (2005).
44. Alexander, M. K. et al. Disrupting Gram-negative bacterial outer membrane biosynthesis through inhibition of the lipopolysaccharide transporter MsbA. *Antimicrob. Agents Chemother*. **62**, e01142-18 (2018).
45. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA* **95**, 5752–5756 (1998).
46. Ladant, D. & Ullmann, A. *Bordatella pertussis* adenylate cyclase: a toxin with multiple talents. *Trends Microbiol*. **7**, 172–176 (1999).
47. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol*. **37**, 911–917 (1959).
48. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc*. **4**, 706–731 (2009).
49. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307–326 (1997).
50. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
51. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr*. **40**, 658–674 (2007).
52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
53. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
54. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
55. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. **44** (W1), W344–W350 (2016).
56. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res*. **44** (W1), W351–W355 (2016).
57. The PyMOL Molecular Graphics System, version 2.0 Schrödinger, LLS.
58. Liu, W., Ishchenko, A. & Cherezov, V. Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat. Protoc*. **9**, 2123–2134 (2014).
59. Barty, A. et al. *Cheetah*: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr*. **47**, 1118–1131 (2014).
60. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst*. **45**, 335–341 (2012).
61. White, T. A. et al. Recent developments in *CrystFEL*. *J. Appl. Crystallogr*. **49**, 680–689 (2016).
62. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des*. **27**, 221–234 (2013).
63. Schrödinger Release 2017-3: Schrödinger Suite 2017-3 Protein Preparation Wizard (New York, 2017).
64. Schrödinger Release 2017-3 (New York, 2017).
65. Shivakumar, D. et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput*. **6**, 1509–1519 (2010).
66. Guo, Z. et al. Probing the α-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. *Chem. Biol. Drug Des*. **75**, 348–359 (2010).
67. Bowers, K. J. et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC '06: Proc. 2006 ACM/IEEE Conference on Supercomputing*, 43–43) (Tampa, FL, 2006).
68. Hankins, J. V., Madsen, J. A., Needham, B. D., Brodbelt, J. S. & Trent, M. S. The outer membrane of Gram-negative bacteria: lipid A isolation and characterization. *Methods Mol. Biol*. **966**, 239–258 (2013).
69. Choi, H. et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods* **8**, 70–73 (2011).
70. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res*. **28**, 235–242 (2000).
71. Ma, G., Zhu, Y., Yu, Z., Ahmad, A. & Zhang, H. High resolution crystal structure of the catalytic domain of MCR-1. *Sci. Rep*. **6**, 39540 (2016).
72. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233–237 (2017).
73. Noland, C. L. et al. Structural insights into lipoprotein N-acylation by *Escherichia coli* apolipoprotein N-acyltransferase. *Proc. Natl Acad. Sci. USA* **114**, E6044–E6053 (2017).
74. Vilar, S., Cozza, G. & Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem*. **8**, 1555–1572 (2008).
75. Owens, T. W. et al. Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550–553 (2019).
76. Ohto, U., Fukase, K., Miyake, K. & Shimizu, T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc. Natl Acad. Sci. USA* **109**, 7421–7426 (2012).

Acknowledgements We thank our Genentech colleagues for their support, in particular, A. Song, I. Kekessie, J. Toms, E. Hecht, C. Peng, A. Liu, P. Smith, A. Estevez, C. Ciferri, H. Ho, E. Castellanos, A. K. Katakam, I. Zilberleyb, M. Reichelt, M.-W. Tan, J. Kiefer, Y. Franke, C. Koth, E. Brown and S. Hymowitz. We thank D. Cawley for antibody generation and Smartox Biotechnology for peptide synthesis. Use of the Linac Coherent Light Source (LCLS) and the Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, are supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. C.G. appreciates support from the SLAC National Accelerator Laboratory and Stanford University through a Panofsky Fellowship. All reagents are available under a materials transfer agreement with Genentech.

Author contributions T.C. and K.R.B. contributed equally to this work. T.C. purified and crystallized PbgA, with support from C.L.N. K.R.B., K.M.S., N.N.N., D. Swem and S.T.R. performed microbiology experiments. Q.L., P.L., E.V., G.H. and W.S. performed proteomics and lipidomics experiments, with support from T.C. J.R. and E.S. performed genomic sequence analyses. D. Sangaraju and S.S.-L. performed metabolomics experiments, with support from T.C. S.P. and M.X. performed in vivo experiments. L.M. and T.D.V. performed molecular biology and protein expression experiments. A.M. performed the LAL assay. D.P.D., M.S.H. and C.G. collected and processed SFX data. T.C. and J.P. determined and refined structures, with input from C.G. B.D.S. performed and analysed molecular dynamics simulations. J.P. proposed the lipid A-binding potential of PbgA-derived peptides and designed variants; N.S. and E.J.H. designed key peptides; M.R.-G. synthesized key peptides; T.C. performed lipid-binding interferometry experiments; K.R.B. and S.T.R. performed bacterial growth inhibition and MIC assays with peptides. T.C., K.R.B., S.T.R. and J.P. wrote the manuscript with input from all authors. J.P. and S.T.R. co-supervised the project and E.J.H., J.P. and S.T.R. are co-senior authors.

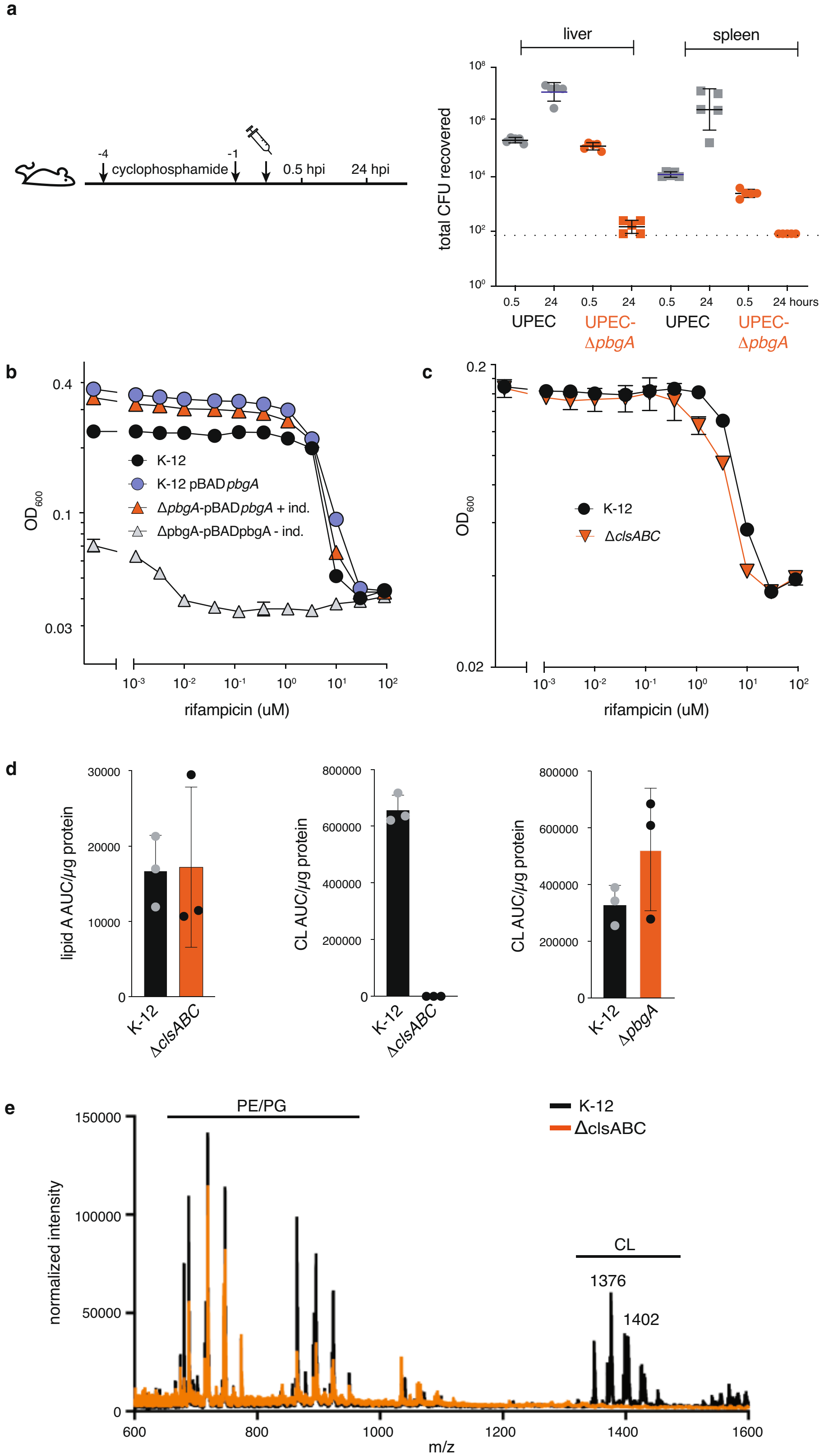
Competing interests All authors, except D.P.D., M.S.H. and C.G., are or were employees of Genentech/Roche.

Additional information

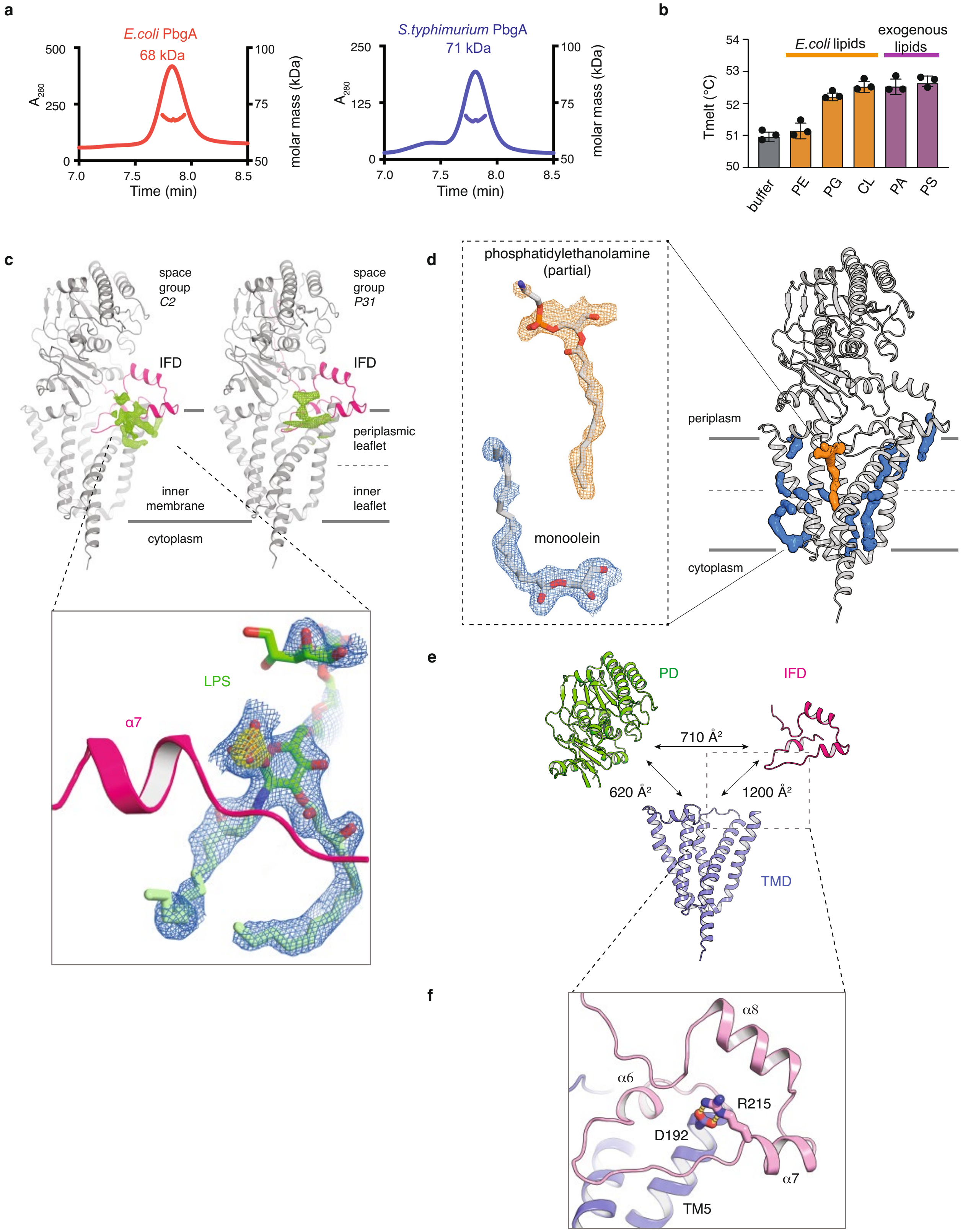
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2597-x.

Correspondence and requests for materials should be addressed to E.J.H., J.P. or S.T.R.

Peer review information *Nature* thanks Bert van den Berg, Russell Bishop, Changjiang Dong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

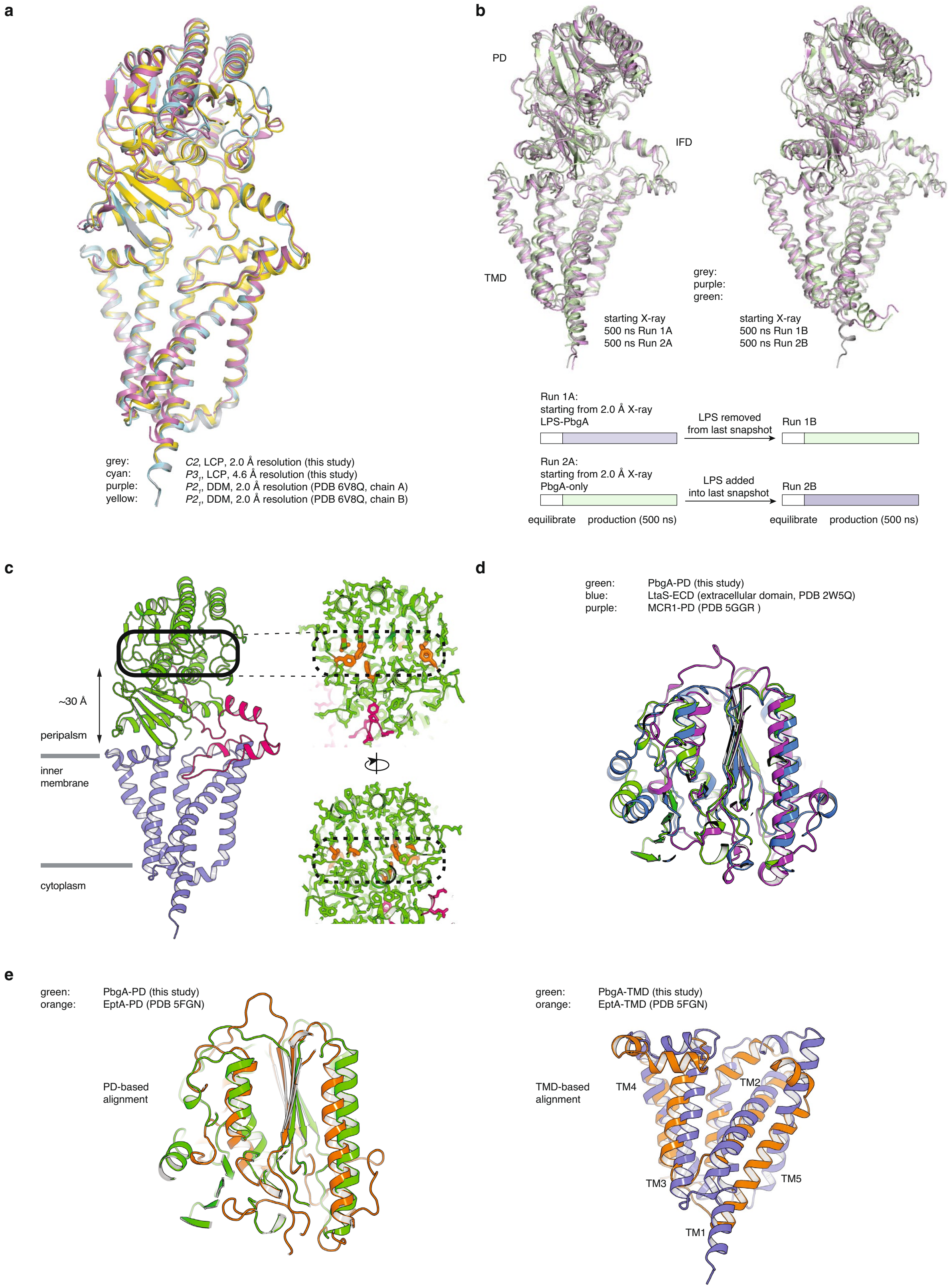


Extended Data Fig. 1 | In vivo and in vitro characterization of *E. coli* Δ*pbgA* and Δ*clsABC* strains. **a**, CFUs recovered from UPEC and UPEC Δ*pbgA* in neutropenic mouse tissues after intravenous injection of BALB/C mice 0.5 and 24 h after injection ($n = 5$ per group). Data are mean ± s.d. with dashed line indicating lower boundary of detection. **b**, Rifampicin sensitivity assay with conditional *E. coli K-12* Δ*pbgA::pBADpbgA* strain. Data are mean ± s.d. for at each rifampicin concentration for $n = 3$ of each strain. **c**, Rifampicin sensitivity assay with *E. coli* K-12 and Δ*clsABC* strains. Data are mean ± s.d. for each rifampicin concentration for $n = 3$ of each strain. **d**, Quantification of lipid A and cardiolipin measured by MALDI–TOF and Qtrap liquid chromatography–tandem mass spectrometry (LC–MS/MS), respectively, normalized to total protein amounts in whole cells (left and middle) or outer membrane vesicles (right). AUC, area under the curve. Data are mean ± s.d. for each strain for $n = 3$ replicates. **e**, MALDI–TOF mass spectrometry analyses detected no cardiolipin in the Δ*clsABC* strain (orange) compared to the *E. coli K-12* strain (black) when analysed under matched conditions. Representative results are shown.



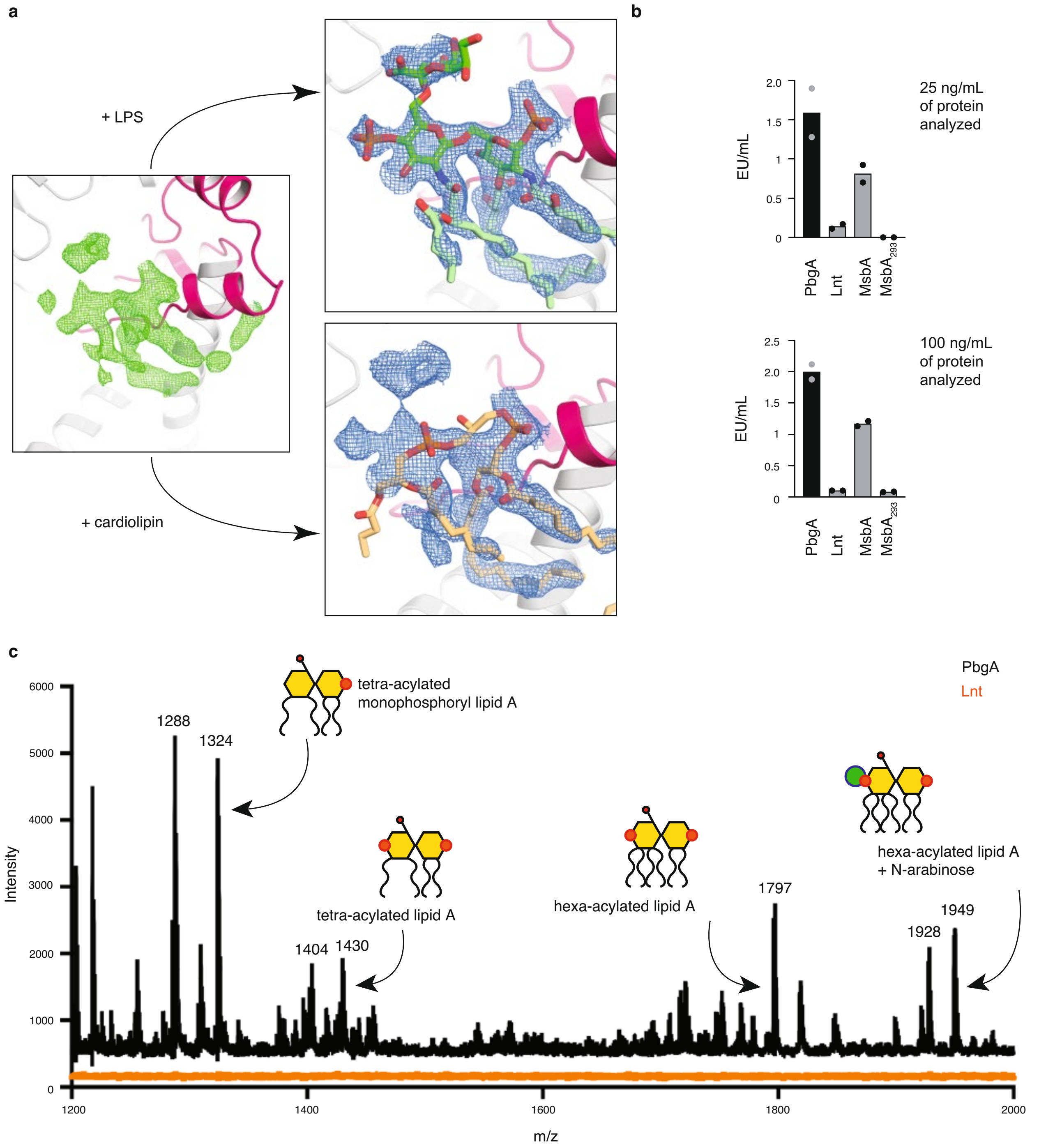
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Biophysical and structural characterization of PbgA. **a**, *E. coli* and *S. typhimurium* PbgA were purified in the mild detergent dodecylmaltoside and analysed by SEC-MALS. **b**, Thermostability of purified *E. coli* PbgA was analysed by differential scanning calorimetry with or without 0.1 mg ml^{-1} lipid supplementation. **c**, Left, from PbgA crystalized in space group *C*2, using data to 2.0 Å, an $F_o - F_c$ map calculated shows bilobal extra electron density along the periplasmic membrane leaflet before the inclusion of LPS into models, 3σ contour. Inset, close-up view of an $F_o - F_c$ map calculated before the inclusion of LPS into the model, rendered at 8σ (yellow) and 2σ (blue), respectively. Final refined coordinates of lipid A are shown for reference. Right, from PbgA crystalized in space group $P3_1$, using data to 4.6 Å, an $F_o - F_c$ map calculated before the inclusion of LPS into the model, contoured at 3σ. **d**, Representative non-protein densities observed surrounding the TMD of PbgA that were assigned as putative phosphatidylethanolamine or monoolein lipids; inset shows $F_o - F_c$ maps calculated before the inclusion of phosphatidylethanolamine or monoolein into the model, 2σ contour (phosphatidylethanolamine, orange; monoolein, blue). **e**, Schematic illustration of the inter-domain surface area contacts within PbgA. **f**, Close-up view highlighting the interaction of the Arg215 side chain with a conserved acidic residue, Asp192 on TM5, which appears to stabilize the IFD–TMD interface.

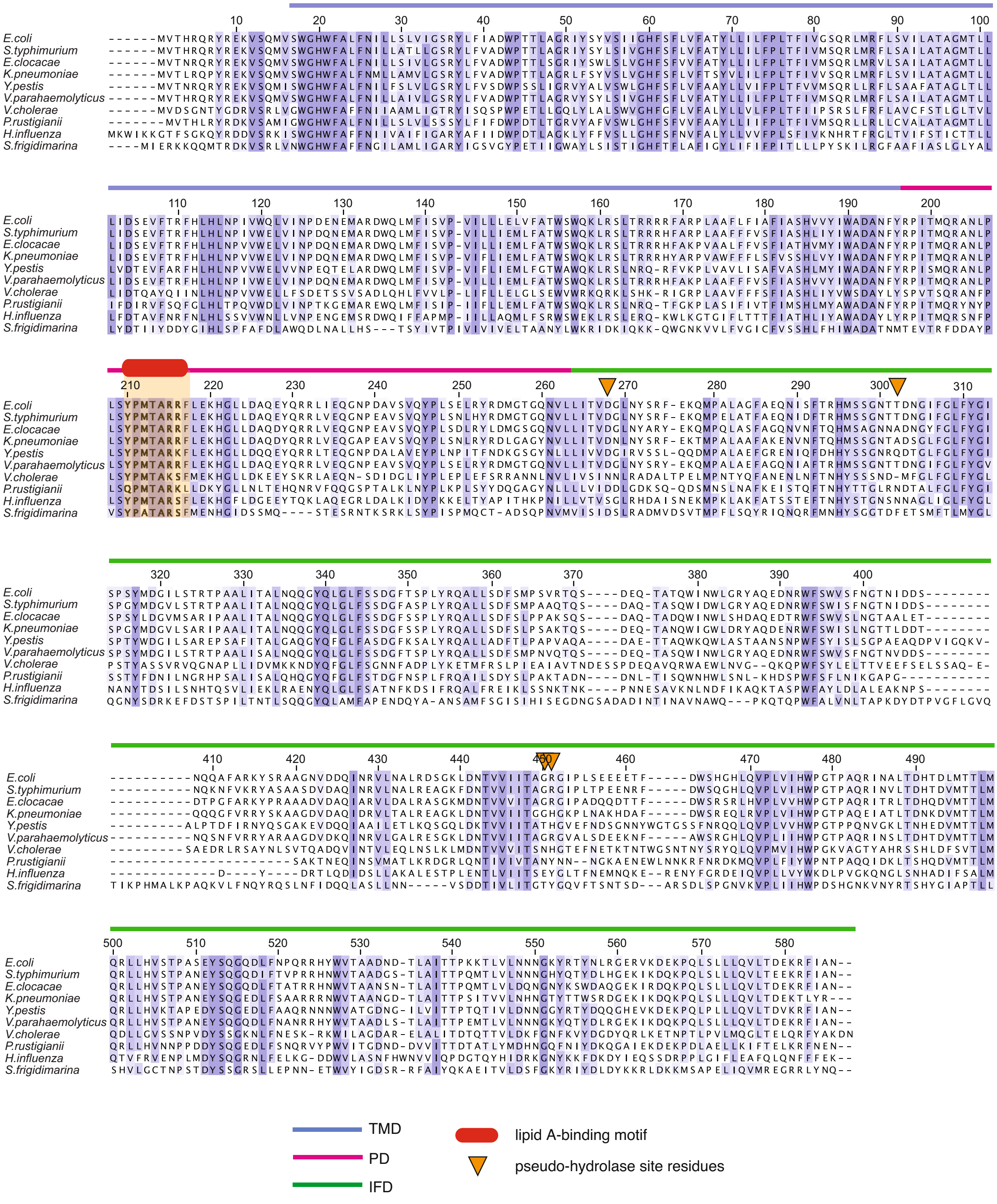


Extended Data Fig. 3 | See next page for caption.

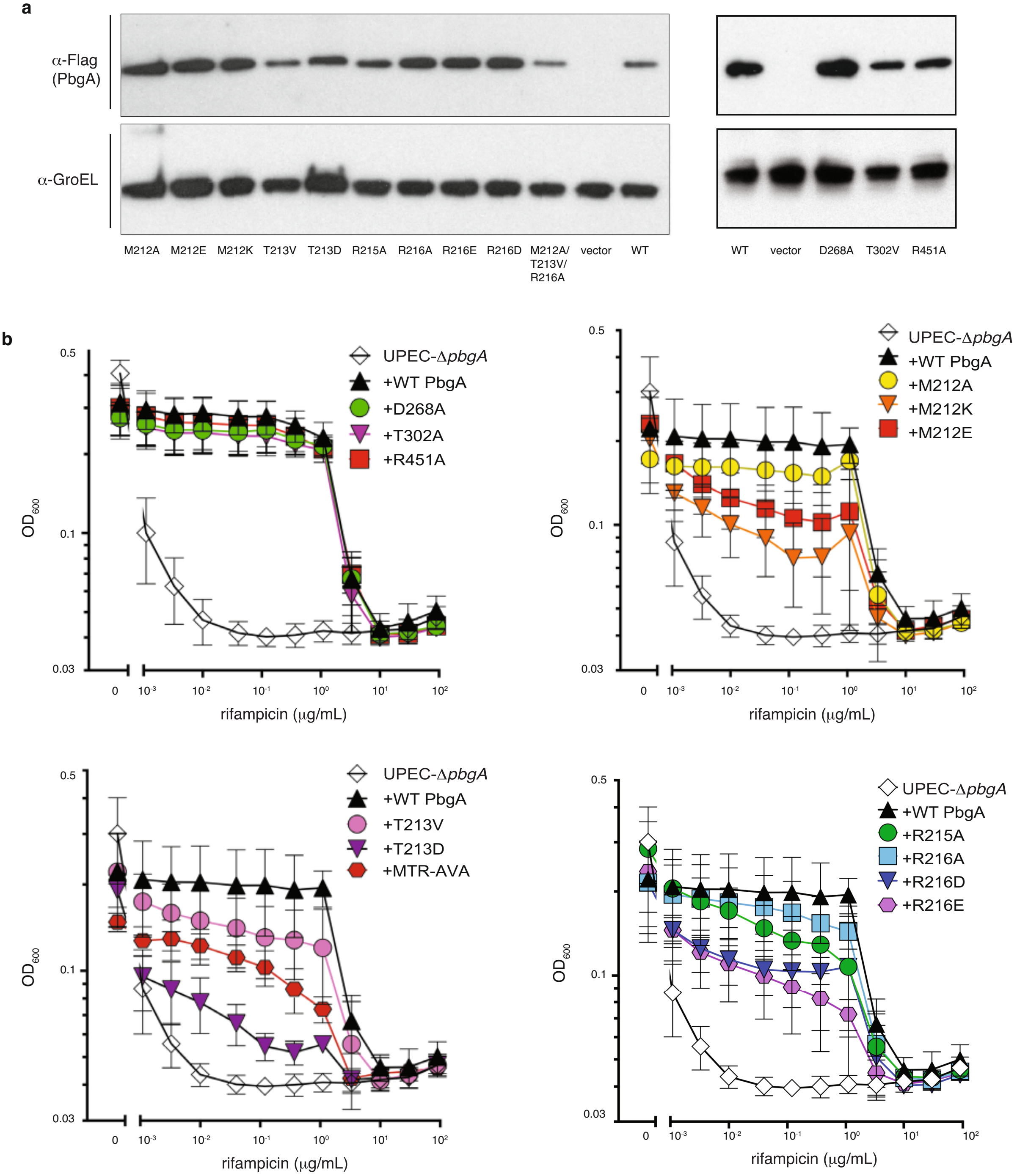
Extended Data Fig. 3 | PbgA structural alignments and molecular dynamics simulations. **a**, Structural superposition of PbgA crystal structures determined in the present study (space group *C*2 and *P*3_1) and both chain A and chain B from PDB code 6V8Q. The overall root mean square deviation for main chain atoms between the most divergent structures is <0.8 Å. **b**, Molecular dynamics study of PbgA, results (top) and experiments (bottom) are summarized by illustration. Simulations were performed following preparation of the 2.0 Å PbgA crystal structure and its placement into a phosphatidylethanolamine: phosphatidylglycerol mixed membrane bilayer, as described in Methods. Top, superimposed are coordinates from the last frames of the four molecular dynamics simulation runs with the starting (non-relaxed) X-ray model to compare the extent of domain movements. **c**, Views of the previously proposed cardiolipin-binding site[8] are shown on the right. Residues proposed to be involved in cardiolipin binding are shown as orange sticks, but are seen here to form an integral part of the hydrophobic protein core; furthermore, the periplasmic domain of PbgA contains no recognizable sequence or structural homology to previously established lipid binding modules[56,70]. **d**, Structure-based alignment of the hydrolase superfamily domains from PbgA (periplasmic domain, green), *S. aureus* LtaS[22] (ECD, blue) and *E. coli* phosphoethanolamine transferase MCR-1[71] (periplasmic domain, purple). **e**, Structure-based alignment of PbgA and EptA isolated periplasmic domains (left) and TMDs (right), respectively.



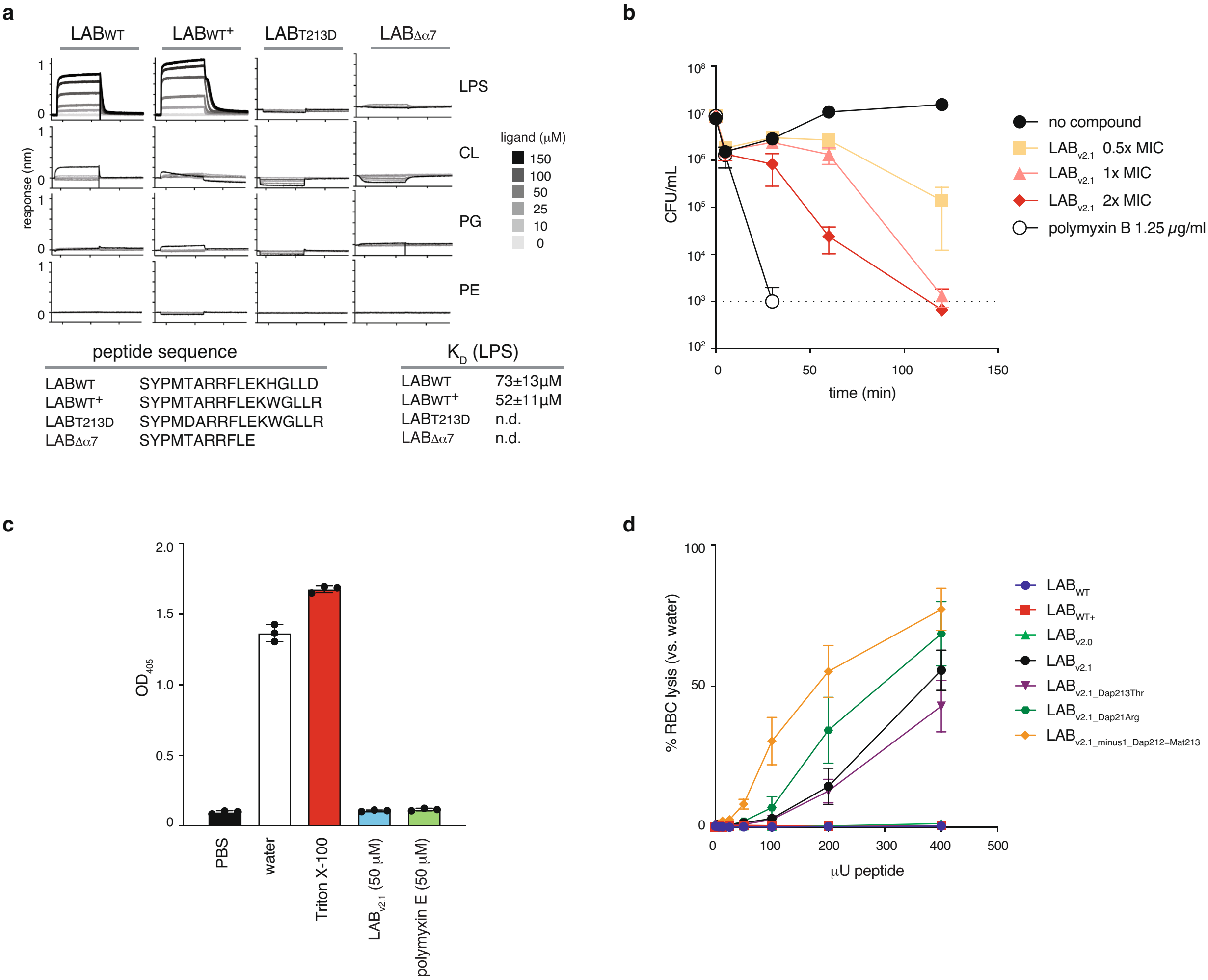
Extended Data Fig. 4 | LPS co-purifies and is bound to PbgA. **a**, Calculated using data to 2.0 Å, an $F_o - F_c$ map near the α7 helix of the IFD (pink) before inclusion of any ligand into refinement, 2σ contour (green). LPS refines well into this electron density whereas cardiolipin does not (see Extended Data Fig. 2c). Modelling and crystallographic refinement was pursued for cardiolipin, phosphatidylethanolamine, phosphatidylglycerol, monoolein and lauryl maltose neopentyl glycol (LMNG) detergent, but all efforts returned unacceptable refinement outcomes and maps. A $2F_o - F_c$ map following the inclusion of LPS into the refinement (blue, 0.8σ contour) is shown for reference. **b**, LPS quantification from proteins purified under matched conditions and subjected to a limulus amebocyte lysate assay. MsbA, the inner membrane LPS transporter from *E. coli*[29,72], was purified from a recombinant *E. coli* expression host and HEK293 cells (MsbA$_{293}$) for comparison. Lnt is an inner membrane protein involved that is not known or expected to bind or transport LPS[73], and was expressed and purified from *E. coli* for comparison. Experiments were run in duplicate at three different protein concentrations with similar results, where duplicate experiment with 25 ng ml^{-1} and 100 ng ml^{-1} protein are shown. **c**, MALDI–TOF mass spectrometry detects various lipid A species from purified PbgA, including an arabinose-modified species (black). No lipid A species were detected from Lnt purified and analysed under matched conditions (orange).



Extended Data Fig. 5 | Sequence alignment of then PbgA homologues. Sequence alignment of ten PbgA sequences from *Enterobacteriaceae* Gram-negative bacteria. Domain boundaries are based on *E. coli* PbgA structure are indicated, including the lipid A-binding motif (red shade) and pseudo-hydrolase active site residues (orange triangles).



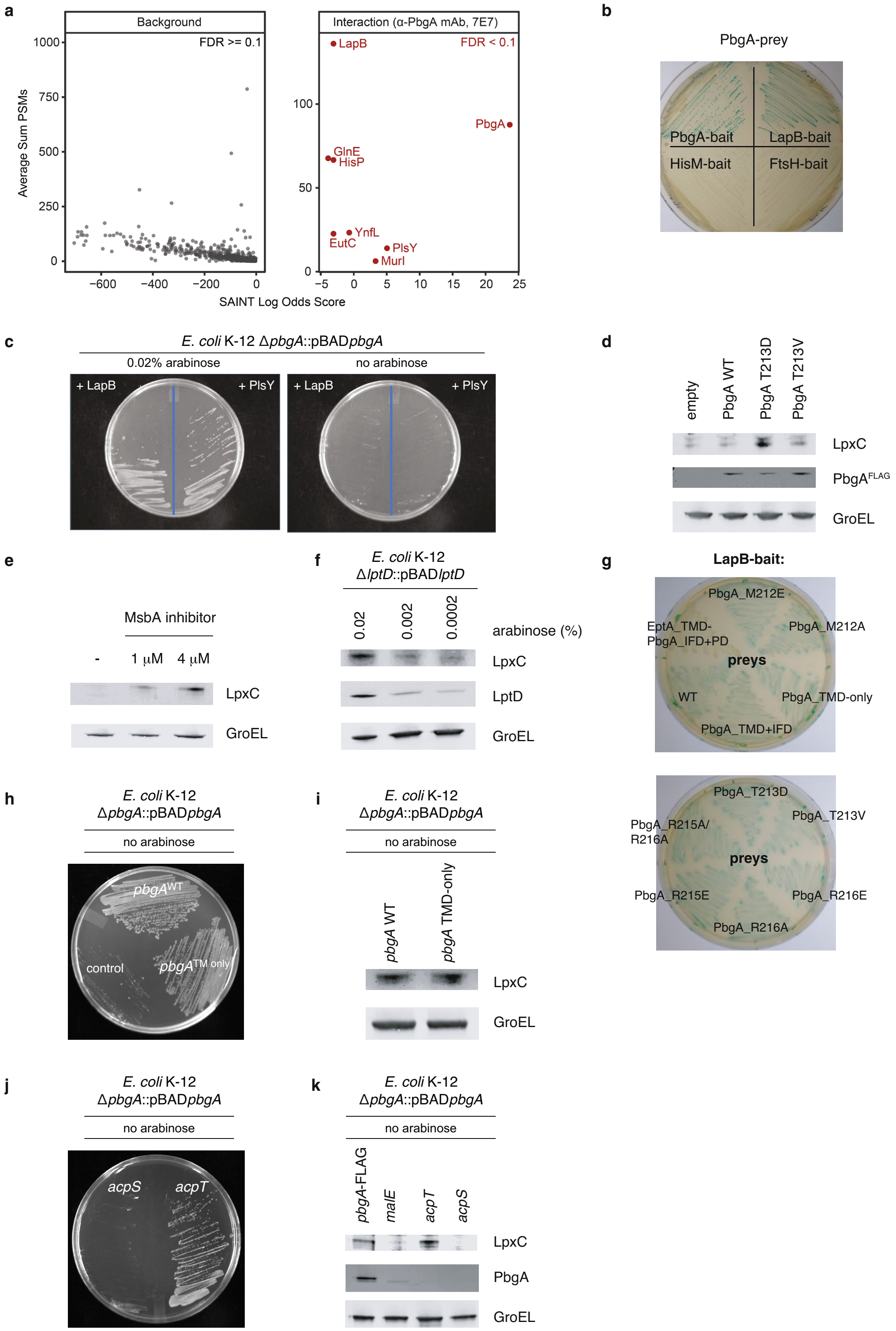
Extended Data Fig. 6 | PbgA mutants and outer membrane permeability. **a**, All UPEC-Δ*pbgA* bacteria tested in the rifampicin sensitivity assay were probed by western blot analysis to confirm PbgA–Flag expression. GroEL was assessed as a loading control. Representative blots for $n = 3$ or more experiments are shown. **b**, Outer membrane permeability of UPEC Δ*pbgA* strains with pBAD*pbgA* plasmids expressing wild-type or mutant *pbgA* assessed by rifampicin sensitivity, where MTR-AVA is the M212A/T213V/R216A PbgA triple mutant. Data are representative and presented as mean ± s.d. for $n = 3$ or more independent cultures. Note, see Extended Data Fig. 2f for a view of the salt-bridge interaction between R215 (IFD) and a conserved TMD acidic residue, D192.



	peptide sequence
LABWT	SYPMTARRFLEKHGLLD
LABWT+	SYPMTARRFLEKWGLLR
LABT213D	SYPMDARRFLEKWGLLR
LABΔα7	SYPMTARRFLE

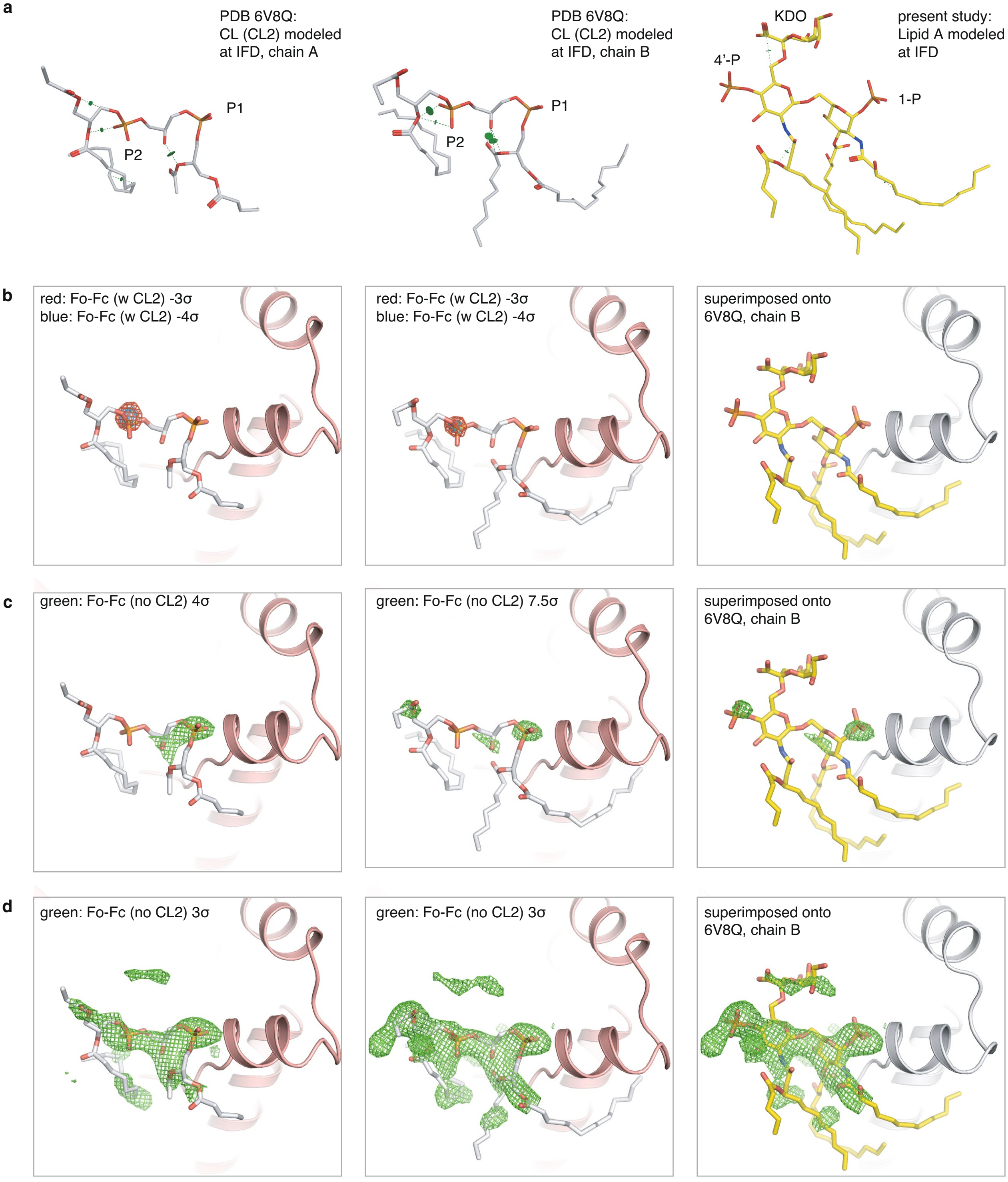
	K_D (LPS)
LABWT	73±13μM
LABWT+	52±11μM
LABT213D	n.d.
LABΔα7	n.d.

Extended Data Fig. 7 | Characterization of PbgA-derived, synthetic LAB peptides. a, Biotinylated LAB peptides were captured and interferometry measurements measured upon presenting peptides to different concentrations of detergent solubilized lipids (LPS, phosphatidylethanolamine, phosphatidylglycerol and cardiolipin). Three independent experiments were performed and data shown are representative. **b**, CFUs of *E. coli* ATCC 25922 measured over time with LAB$_{v2.1}$ and polymyxin B. Data are mean ± s.d. for *n* = 3 independent cultures. **c**, A red blood cell (RBC) lysis assay evaluated after 18 h in the presence of indicated compounds (Methods). Data are mean ± s.d. (*n* = 3) for each compound tested. **d**, A RBC lysis assay comparing LAB$_{v2.1}$ precursors (LAB$_{WT}$, LAB$_{WT+}$, LAB$_{v2.0}$) and LAB$_{v2.1}$ analogues designed, based on the LPS–PbgA crystal structure, to disrupt specific interactions of lipid A (LAB$_{v2.1\_Dap213Thr}$, LAB$_{v2.1\_Dap213Arg}$, LAB$_{v2.1\_Dap212\text{-}Met213}$). Data are mean ± s.d. for *n* = 3 independent assay of each compound at each concentration.



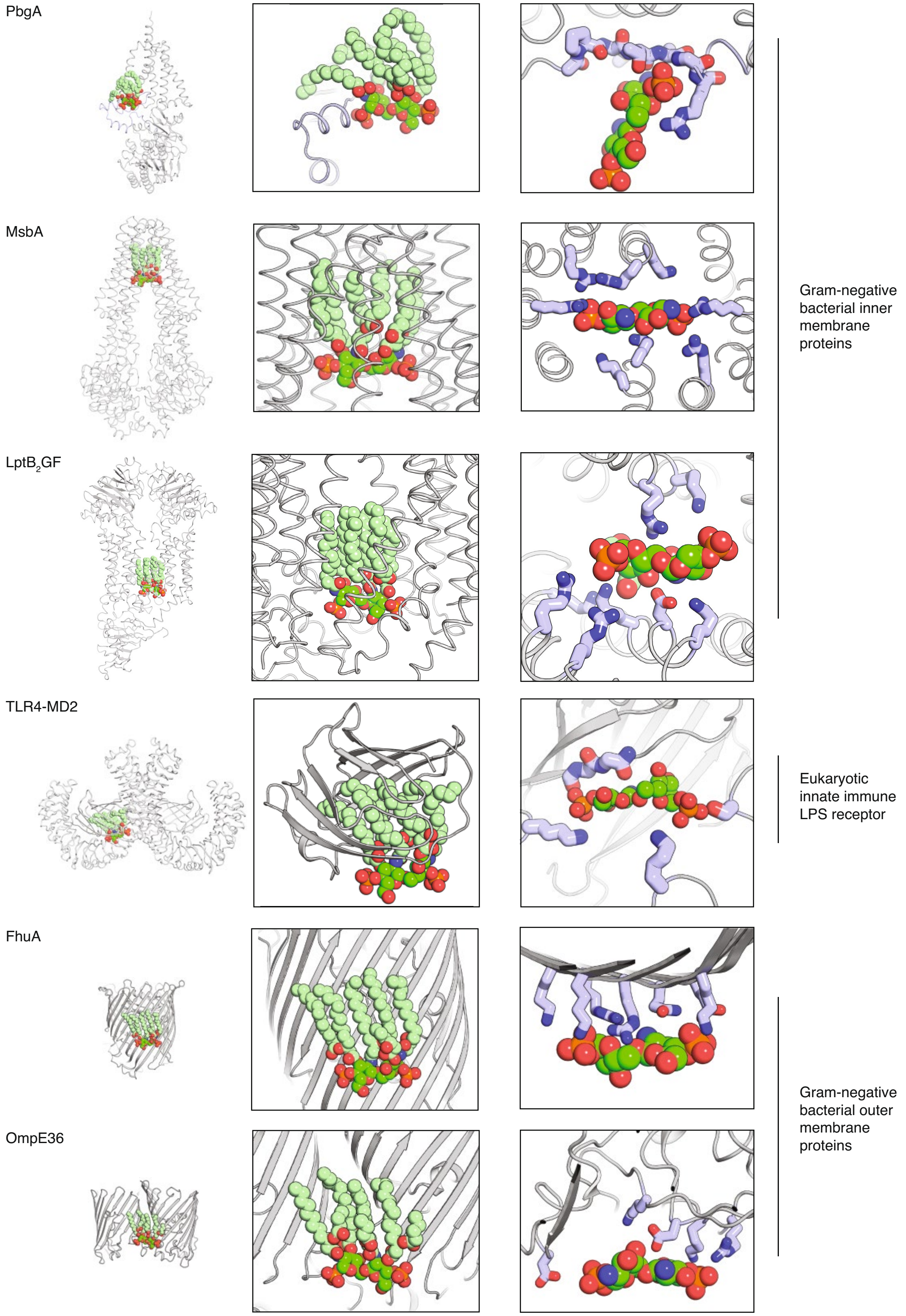
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | PbgA interacts with LapB to regulate LpxC stability. **a**, Proteins identified by mass spectrometry following co-immunoprecipitation of endogenous PbgA using the anti-PbgA monoclonal antibody 7E7 (*n* = 3 independent experiments). Hits were classified based on abundance (sum of PSMs) and enrichment in PbgA IPs compared to control purifications (SAINT logOddsScore: anti-PbgA monoclonal antibody 7E7 versus anti-gp120). Identified proteins with a Bayesian FDR <10% are highlighted in red. **b**, Bacterial two-hybrid system using PbgA-prey and different bait proteins in *E. coli* cells. Interacting proteins lead to blue colonies on agar plates containing X-gal, whereas non-interacting proteins produce white colonies. A representative agar plate is shown (*n* = 3) and activity was confirmed in broth cultures. **c**, Growth of a conditional *E. coli* K-12 Δ*pbgA*::pBAD-*pbgA* after depletion of PbgA in the presence of a IPTG-inducible plasmid expressing wild-type *lapB* or *plsY* (Methods) demonstrates that *lapB* expression does not rescue growth after PbgA depletion. Representative plates are shown and growth assay was repeated three or more times. **d**, Cell lysates prepared from overnight streaks of *E. coli* K-12 with pBAD*pbgA* wild-type or mutant plasmids were probed with anti-LpxC, anti-PbgA and anti-GroEL antibodies (Methods), indicating that disturbing the LPS–PbgA interaction interface leads to LpxC stabilization. Representative blots from *n* = 3 biological replicates are shown. **e**, Western blot analysis of LpxC after treatment with 1 μM (2× MIC) or 4 μM (8× MIC) of the small molecule MsbA inhibitor G′913, indicating that selective inhibition of MsbA[29,44] and LPS transport impacts LpxC levels; GroEL is the loading control and a representative experiment (*n* = 3 independent experiments) is shown. **f**, *E. coli* K-12 Δ*lptD*::pBAD*lptD* lysates prepared from cells grown in indicated concentration of arabinose were probed with anti-LpxC, anti-LptD and anti-GroEL antibodies (Methods). Representative blots from *n* = 3 biological replicates are shown. **g**, Bacterial two-hybrid assays using LapB-bait (pUT18-*lapB*) and indicated PbgA-mutant prey constructs (pKT25-*pbgA*) in *E. coli* DHM1 cells were performed (Methods). Interacting proteins lead to blue colonies, whereas non-interacting proteins produce white colonies. Note that EptA$^{\mathrm{TM}}$–PbgA$^{\mathrm{IFD+PD}}$ is a chimeric construct in which the TMD of PbgA has been replaced with the TMD region from EptA[23]. Representative plates from *n* = 3 culture streaks are shown. **h**, Growth of conditional PbgA strain (*E. coli* Δ*pbgA*::pBAD*pbgA*) in the absence of arabinose inducer complemented with, clockwise from the top of plate, wild-type *pbgA* (PbgA$^{\mathrm{WT}}$), *pbgA* encoding only the TMD (PbgA$^{\mathrm{TM\ only}}$), or a negative control (*malE*) on plasmids. A representative plate (*n* = 3) is shown. **i**. Cell lysates of the conditional *pbgA* strain (*E. coli* Δ*pbgA*::pBAD*pbgA*) in the absence of arabinose inducer complemented with wild-type *pbgA* or *pbgA* encoding only the TMD were probed with anti-LpxC antibody (Methods). A representative blot for *n* = 3 independent experiments is shown. **j**, Plasmids encoding *acpT* (right side of plate) or *acpS* (left side of plate) in conditional-*pbgA* strain grown in the absence of the pBAD*pbgA* inducer arabinose, with 0.1 mM IPTG at 30 °C. A representative growth plate (*n* = 3) was imaged. **k**, Cultures with plasmids expressing *pbgA*, *acpT*, *acpS*, *or malE* (control) were shifted to no arabinose/plus IPTG if necessary to deplete PbgA (Methods). A representative blot from at least *n* = 3 biological replicates is shown.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | A previous PbgA crystal structure reported to have cardiolipin bound at the IFD is, instead, more consistent with bound lipid A. **a**, At the inner membrane–periplasmic interface that we term the IFD: cardiolipin (named CL2)[7] from chain A (left) and chain B (middle) of PDB 6V8Q are shown in stick representation; PbgA is removed for clarity. Similarly, lipid A is shown in stick representation taken from the high-resolution crystal structure presented in this work (right). Molecular clashes calculated using the MOE software[74] indicate high-energy atomic distance and poor geometry (green lines) in both chains A and B from PDB 6V8Q. The extent of the intramolecular clash is indicated by the relative size of the green circle. **b**, An $F_o - F_c$ map calculated using coordinates and structure factors from PDB 6V8Q chain A (left) and chain B (middle) shows a strong negative peak (−3σ, red mesh; −4σ, blue mesh) on the assigned modelled P2 phosphate position of the CL2 ligand. Right, the LPS–PbgA complex determined in this work is superimposed onto chain B of PDB 6V8Q for reference, with no further adjustments. **c**, An $F_o - F_c$ map calculated using coordinates and structure factors from PDB 6V8Q, with CL2 omitted from the calculation, shows strong positive peaks (4σ and 7.5σ for chain A and B, respectively; green mesh), which, in both cases, appear better described by the LPS–PbgA complex structure determined in this work. Shown (right) is the LPS–PbgA complex superimposed onto chain B of PDB 6V8Q with no further adjustments. **d**, The same $F_o - F_c$ map calculation as in **c**, only contoured to 3σ (green mesh). As seen on the right, when superimposed onto chain B of 6V8Q, the proximal 1-phospho-GlcNAc group of lipid A in our LPS–PbgA structure appears especially well accounted for by positive density peaks, and density consistent with a KDO sugar head group is also observed; and similar conclusions are reaching upon inspection of superposition onto chain A of 6V8Q (not shown).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Comparison of LPS coordination in PbgA to known selective and passive LPS-binding proteins. PbgA (this study), MsbA (PDB code 6BPP), a selective LPS transporter[29,72], LptB$_2$FG (PDB code 6MHU), a selective LPS[33,75], and TLR4-MD2 (PDB code 3VQ2), a high-affinity LPS receptor[32,76], represent the examples of selective LPS-binding proteins with known structures. In these latter cases, the hydrophobic acyl chains of lipid A are increased and the bivalent and polar nature are the lipid A head group is exploited. Furthermore, note that Arg216 of PbgA, shown in stick representation, does not appear essential for binding LPS in vivo (see Fig. 3c). In addition, FhuA (PDB code 2FCP), found with LPS complexed along the outer leaflet region of this outer membrane protein barrel[34], and OmpE36 (PDB code 5FVN), which has also revealed numerous LPS contacts along the barrel[35], are shown for completeness and comparison. Notably, analogous to MsbA, LptB2FG and TLR4, hydrophobic and aromatic side chains make several contacts in FhuA and OmpE36 with the acyl chains of lipid A (not shown for clarity) and polar and basic side chains coordinate the bivalent lipid A head group. In all cases, the lipid A coordination schemes are distinct from what is observed in the LPS–PbgA complex (also see Fig. 3).

nature research

Corresponding author(s): Emily Hanan, Jian Payandeh, Steven Rutherford

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All equipment specifications and experimental parameters have been detailed.
Data analysis	All software and data analyses methods have been described and references appropriately cited. The following software (version numbers) were used: GSNAP (2013-10-10), GenomicRanges (1.34.0), GenomicAlignments (1.18.1), VariantTools (1.24.0), gmapR (1.24.2), PRISM (8.3.1), NIS-Elements AR (4.3), Coot (0.89), ASTRA (6), Phaser (2.8), PHENIX (1.12-2829), PyMOL 2.0.7 (The PyMOL Molecular Graphics System, Schrödinger, LLC), Cheetah (2017.3), CrystFEL (0.6.2), MOE (2019.0101), Maestro Schrodinger (2017-3), Compound Discoverer (2.1.0.401), Dali server (http://ekhidna2.biocenter.helsinki.fi/dali/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All relevant data sets are available in public databases and are available upon request. Structural data are deposited in the protein data bank (PDB) under accession number ABCD. All mass spectrometry RAW files were uploaded to the MassIVE data repository, accessible by the identifier MSV000083754, and can be downloaded from ftp://MSV000083754@massive.ucsd.edu. DNA sequencing data were deposited at NCBI under BioProject PRJNA541088, BioSample SAMN11572257, experiment SRX5788703, run SRR9010525. The E. coli CFT073 reference genome was deposited at NCBI under BioProject PRJNA624646, BioSample SAMN14575425, Accession CP051263.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples sizes were not statistically predetermined. Mouse numbers (n=8 per tested strain for thigh infection model and n=5 per tested strain for IV infection model) were chosen to allow for replicative observations while considering ethical of animal use.
Data exclusions	No data were excluded.
Replication	Experiments were replicated as described in the Figure legends and methods. All gel images shown are representative of replicates.
Randomization	Randomization is not relevant to the growth, structural, or biochemical experiments described in this work.
Blinding	Data were not blinded. Blinding is not relevant to the growth, structural, or biochemical experiments described in this work as subjective analyses were not used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-PbgA (Genentech), anti-GroEL (Enzo), anti-FLAG (Cell Signaling Technologies), anti-LpxC (LSBio), anti-LptD (Genentech), anti-mouse (LI-COR), anti-rabbit (LI-COR), anti-human (LI-COR)
Validation	Anti-PbgA antibodies were generated for this work and are described in the manuscript. Anti-LptD antibodies were described and validated in Storek et al. eLife 2019;8:e46258. Information about anti-GroEL (www.enzolifesciences.com), anti-FLAG (www.cellsignal.com), and the LI-COR (www.licor.com) secondary antibodies is available at the indicated websites. Information about anti-LpxC is available at the manufacture website (www.lsbio.com) with additional Western blot validation in the Supplementary Information. All Western blots, biolayer interferometry, and immunoprecipitations were performed with appropriate controls.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Expi293 (human embryonic kidney cells), purchased from Thermo Fisher Scientific (A14527)
Authentication	The cell lines were not authenticated.
Mycoplasma contamination	The cells tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For thigh infection model, 6 week old CD1 mice (Charles River Laboratories) were used (n=8 animals per tested strain; total n=16). For IV infection model, 7 week old A/J mice (Jackson Laboratory) were used (n=5 animals per tested strain; total n=10).
Wild animals	No wild animals were used in these studies.
Field-collected samples	No field-collected samples were used in these studies.
Ethics oversight	All mice used in the in vivo studies were housed and maintained at Genentech in accordance with American Association of Laboratory Animal Care guidelines. All experimental studies were conducted under protocols approved by the Institutional Animal Care and Use Committee of Genentech Lab Animal Research in an Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC)-accredited facility in accordance with the Guide for the Care and Use of Laboratory Animals and applicable laws and regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Whole blood samples were collected from human volunteers for red blood cell lysis assays. No personal or medical history was specified, provided, or collected for volunteers.
Recruitment	Samples were collected from volunteers.
Ethics oversight	Collection of blood samples was through the Genentech Samples for Science Program and carried out under protocols approved by the Western Institutional Review Board (protocol number CEHS-CP 307.2, IRB tracking number 20080040)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Corrections & amendments

Publisher Correction: Extensive signal integration by the phytohormone protein network

https://doi.org/10.1038/s41586-020-2585-1

Correction to: *Nature* https://doi.org/10.1038/s41586-020-2460-0

Melina Altmann, Stefan Altmann, Patricia A. Rodriguez, Benjamin Weller, Lena Elorduy Vergara, Julius Palme, Nora Marín-de la Rosa, Mayra Sauer, Marion Wenig, José Antonio Villaécija-Aguilar, Jennifer Sales, Chung-Wen Lin, Ramakrishnan Pandiarajan, Veronika Young, Alexandra Strobel, Lisa Gross, Samy Carbonnel, Karl G. Kugler, Antoni Garcia-Molina, George W. Bassel, Claudia Falter, Klaus F. X. Mayer, Caroline Gutjahr, A. Corina Vlot, Erwin Grill & Pascal Falter-Braun

In Fig. 1m of this Article, owing to an error in the production process, the top circle on the right—illustrating type II pathway contact points—was coloured all green instead of half red and half green. This error has been corrected online.

Publisher Correction: Discovery and engineering of colchicine alkaloid biosynthesis

https://doi.org/10.1038/s41586-020-2606-0

Correction to: *Nature* https://doi.org/10.1038/s41586-020-2546-8

Ryan S. Nett, Warren Lau & Elizabeth S. Sattely

In Fig. 4a of this Article, owing to an error in the production process, the enzyme involved in the final step of the scheme (describing the synthesis of compound **10** from **9**) was labelled 'CYP71FB' instead of 'CYP71FB1'. The original Article has been corrected online.

Retraction Note: Exploring the quantum speed limit with computer games

https://doi.org/10.1038/s41586-020-2515-2

Retraction to: *Nature* https://doi.org/10.1038/nature17620

Jens Jakob W. H. Sørensen, Mads Kock Pedersen, Michael Munch, Pinja Haikka, Jesper Halkjær Jensen, Tilo Planke, Morten Ginnerup Andreasen, Miroslav Gajdacz, Klaus Mølmer, Andreas Lieberoth & Jacob F. Sherson

We, the authors, are regretfully retracting this Article owing to an error in our computer code that means the quantitative results reported are not valid. We thank A. Grønlund and D. Sels, whose independent efforts[1,2] pointed to potential problems with our optimization algorithm. The error was identified by A. Grønlund, who has provided a detailed account[3] of the error and its effect on the quantitative results in our Article. For more recent and comprehensive explorations of the performance differences between player-seeded and randomly seeded algorithms, we refer to our recent work[4].

1. Sels, D. Stochastic gradient ascent outperforms gamers in the Quantum Moves game. *Phys. Rev. A* **97**, 040302 (2018)
2. Grønlund, A. Algorithms clearly beat gamers at Quantum Moves: a verification. Preprint at https://arXiv.org/abs/1904.01008 (2019).
3. Grønlund, A. Explaining the poor performance of the KASS algorithm implementation. Preprint at https://arXiv.org/abs/2003.05808 (2020).
4. Jensen, J. H. M. et al. Crowdsourcing human common sense for quantum control. Preprint at https://arXiv.org/abs/2004.03296 (2020).

Work

Your story Send your careers story to: naturecareerseditor @nature.com

MICHAEL SHORT/BLOOMBERG/GETTY

Running a brewing business combines scientific techniques with soft skills honed during a PhD.

THE BREWS AND BAKES THAT FORGED CAREER PATHS

How yeast has helped these scientists experiment with their careers. **By Nikki Forrester**

Many scientists start hobbies to take their minds off research and to connect with people outside academia. Some make these pastimes their careers. *Nature* spoke to four researchers who turned their brewing and fermentation hobbies into business ventures. The scientists – all at different stages in their careers and with varying connections to academic institutions – share their insights.

ANDREW RHODES
MAKING USE OF PHD EXPERIENCE

I started brewing kombucha tea in 2016 after finishing an internship at NASA's Johnson Space Centre in Houston, Texas. One of my colleagues was home-brewing the probiotic-rich, semi-sweet, semi-tangy, fermented beverage, and I thought it was delicious and super fun to brew. I brought home a symbiotic culture of bacteria and yeast, or SCOBY, used to produce the tea, and began brewing kombucha as a fun gig during a period when my PhD research at West Virginia University in Morgantown wasn't going as smoothly as I had hoped. It released my mind from research and gave me an opportunity to work on something that was operating correctly – it gave me a moment of success.

Towards the end of my PhD, I started looking into career opportunities outside university. Along with a passion for teaching, I've always wanted to become an entrepreneur. In 2018, my wife and I applied to the West Virginia Business Plan Competition, which is a contest for university students in the state to develop a business plan and receive funding for their idea. We wrote a plan for a kombucha brewery, or kombuchery, performed a feasibility study to determine whether our product was viable in our market, and I gave a 15-minute pitch in front of an audience and judges. In April 2019, we received US$12,000 to start our business.

In summer 2019, my wife and I found a location for the kombuchery and started purchasing equipment. I graduated the following December, so I worked on the business and my PhD together for a while. It was difficult writing a dissertation and then brewing kombucha late into the night. But by having two directions – research and the kombuchery – I maintained my excitement for both.

With kombucha, I'm juggling multiple brewing cycles and different flavours, while also dealing with accounting, distribution and sales. It's the same as being in a PhD programme, in which you're writing a journal paper, teaching a course and taking a class. Graduate studies teach you time management. I also learnt public speaking through teaching and presenting at conferences, which helped in the business competition because I felt confident in front of the judges.

A few months ago, I secured a full-time job as a teaching assistant professor in aerospace engineering at West Virginia University. It's a once-in-a-lifetime opportunity, so I had to take it. I'm not quitting the kombucha business, so I'll be juggling university and brewing again.

Andrew Rhodes is an aerospace engineer and founder of the Neighborhood Kombuchery in Morgantown, West Virginia.


JOEL WOLPERT

Andrew Rhodes and his wife, Carissa Herman, run a kombucha brewery in West Virginia.

RICHARD PREISS
BLENDING RESEARCH AND BREWING

I became curious about yeast and beer in 2012, when one of my undergraduate housemates at the University of Guelph in Canada started home-brewing. If you are scientifically minded, home-brewing gives you a chance to practically apply knowledge about biology, chemistry and physics, and at the end of the process, you end up with beer. As a microbiologist, I'm used to thinking about tiny organisms – but a lot of people don't think about yeast's role in beer, because it's less tangible than hops or malt. A lot of the flavour in beer comes from yeast. You have to use a lager yeast to make a lager, for example. The flavour of a saison beer is just pure yeast expression – you're letting the yeast take centre stage.

Back in 2012, I had access to a research laboratory and started storing some of the yeast I was using for home-brewing in a cryogenic freezer for long-term safe-keeping and periodic retrieval. Another researcher in the lab noticed the yeast and suggested we approach some of the local breweries to see if we could trade the yeast we grew for beer. The breweries were excited about potentially having a local supplier for yeast instead of importing it into Canada. We also worked with some local brewers to test and share Ontario wild yeasts. A few of these brewers mentioned we could start a business instead of offering our yeast in exchange for beer.

In 2015, we founded Escarpment Laboratories in Guelph to supply liquid yeast cultures to craft and home brewers. Now, we have a core list of about 30 yeasts or blends that we sell – each of which has its own flavour and chemistry. Our frozen collection has about 1,500 strains of yeast and other microbes.

Part of what fuels me is that I get to participate in research all the time. We do research internally for our product development and we work with academics at several institutions, including the University of Guelph and the University of Waterloo. We focus on understanding the natural diversity in flavours and functions of beer yeasts.

We often conduct experiments in which we put up to 50 yeast strains in the same environment and see which aroma and flavour traits are expressed. We're also sequencing the genomes of these yeast strains to understand which genetic traits might underlie flavour production or properties such as alcohol tolerance and aroma production. Once we have that fundamental knowledge, we can start getting creative about customizing, hybridizing and modifying yeast strains.

Being a scientist prepares people for life in business, especially entrepreneurship, because science and business both involve experimentation and failure. You have to think in an agile manner, change plans on the fly and be creative. As a scientist, you learn how to deal with failure because sometimes 80% of your experiments don't work.

Richard Preiss is a microbiologist and co-founder of Escarpment Laboratories in Guelph, Canada.

J. NIKOL JACKSON-BECKHAM
ACADEMIA WITHOUT THE ACADEMY

I got into craft beer in the late 1990s, when I was an undergraduate at Virginia Polytechnic Institute and State University in Blacksburg. I started making it for personal consumption during a master's programme in communication studies at San Diego State University in California. The craft-beer scene was massive in that area. During my PhD, I worked as a manager at several stores that sold supplies for home-brewing. I kept thinking, if I'm this into beer and I'm going to graduate school, why don't I allow these two worlds to overlap?

My PhD dissertation was about beer, how its value was formed and manipulated in the US brewing industry from prohibition in the 1920s onwards. Although I enjoyed my PhD at the University of North Carolina at Chapel Hill, I kept thinking about equity and inclusion in the beer world. I've always been curious about the ethnic and gender disparities in the industry in terms of who drinks craft beer and who makes craft beer.

According to a survey conducted by the Brewers Association (BA), craft breweries in the United States are overwhelmingly owned by white people – people of colour own just a few per cent. There's also not much ethnic or gender diversity among brewers. Women and people of colour tend to be in front-of-house jobs such as bartending and serving. Even though that's not what my dissertation was about, I still wanted an outlet to explore those questions, so I started blogging and making visual art on my personal website.

After several years of working full-time in academia, a friend informed me that the BA was looking for a diversity ambassador. It was a part-time, contract position to conduct industry research, write educational materials, give seminars and work with the diversity committee to create programmes, including a Diversity and Inclusion Events Grant programme and a mentorship programme for under-represented people who want to get involved in the beer industry. After I became the BA's first diversity ambassador in April 2018, breweries started asking me about individual consultations. A few months later, I launched Crafted For All, a platform for my consulting work with the BA, individual breweries and other brewing associations.

In mid-2020, I left Randolph College in Lynchburg, Virginia, where I was a professor of communication studies. Last September, I started the non-profit organization Craft × EDU in Richmond, Virginia. It champions inclusion, equity and justice in the craft-brewing community through education and professional development. We have a few core programmes, including opportunity fairs at which the craft-brewing industry is introduced to people from under-represented communities who are seeking employment.

When people ask me what it's like to no longer be an academic, I always say I'm definitely an academic, I just left the academy. A tenure-track job has three conventional roles: teaching, research and service. I teach through giving keynote talks and seminars. I do research by collecting data for the BA and leading data-driven projects and surveys as a consultant. And I do service as an executive director of a non-profit organization. All the research skills and analyses are still there, but when people look at my work now, they make decisions and implement recommendations in their workplaces.

J. Nikol Jackson-Beckham is a communication studies scholar, founder of Crafted for All and executive director of Craft x EDU in Richmond, Virginia.

CRAFTED FOR ALL, LLC

J. Nikol Jackson-Beckham has founded two consultancies to improve diversity in brewing.

ANDREW STRANG

SOURDOUGH OPENED A NEW DOOR

I never really had a career plan, but knew that doing a PhD in physics at Imperial College London would leave a lot of doors open to me and I enjoyed doing research in interference optics.

I got very interested in making bread during my PhD and started selling it to friends and delivering it by bike. Towards the end of my programme, I started looking into careers in physics as well as bakery businesses in London. Those bakeries inspired me to start my own.

In 2017, I opened the Bread By Bike bakery in London with a few friends. We launched the business on a shoestring, with a small campaign on the crowdfunding platform Kickstarter and without borrowing money from a bank. It was a real do-it-yourself project and the learning curve was steep. Now, there are 21 people working at Bread By Bike and we've upgraded to an electric bike so we can carry about 80 kilograms of bread on deliveries to restaurants, cafes and bars. We decided to offer a home delivery service when people went into lockdown in London in March. A friend of mine wrote software to manage orders and delivery routes. We've been super busy and will probably stick with home deliveries after the coronavirus situation eases. My job has evolved from baking, delivering and cleaning into a managerial role.

Although the direct skills I was using in the lab are not much use for running a bakery, some of the skills I developed doing scientific research are useful in bread production. Bread is an amazing thing. There's magic in every step of the process. Classic sourdough is just flour, water and salt, fermented without commercial yeast. But there are so many ways that making naturally fermented breads can go wrong, which is what's so addictive about it.

Even though I bake the same breads every day, they react completely differently because microbial activity and some of the chemistry depends on the temperature and season. If you made sourdough in the lab, you'd be able to control those parameters and have a very consistent product, but we don't have that. To produce a consistent product, I need to balance a scientific approach of trying to understand why things are happening and an intuitive understanding of how a dough is going to behave.

If you're starting a business, it's important to be honest with yourself about what you want from it and why. Do you want to make millions of pounds? Do you want to be a cornerstone of the community? Once you understand what you want from your business, you can work out how to achieve that step by step.

The key thing for me was the energy of a bakery – the activities, the sounds, the heat, the products coming out of the oven and chucked onto the rack. It's extremely dynamic. I was inspired to create something like that for myself, and although my door to an academic career in physics has probably closed, it's been a great journey making the bakery happen.

Andrew Strang is a physicist and founder of Bread By Bike in London, UK.

Interviews by Nikki Forrester.
These interviews have been edited for length and clarity.

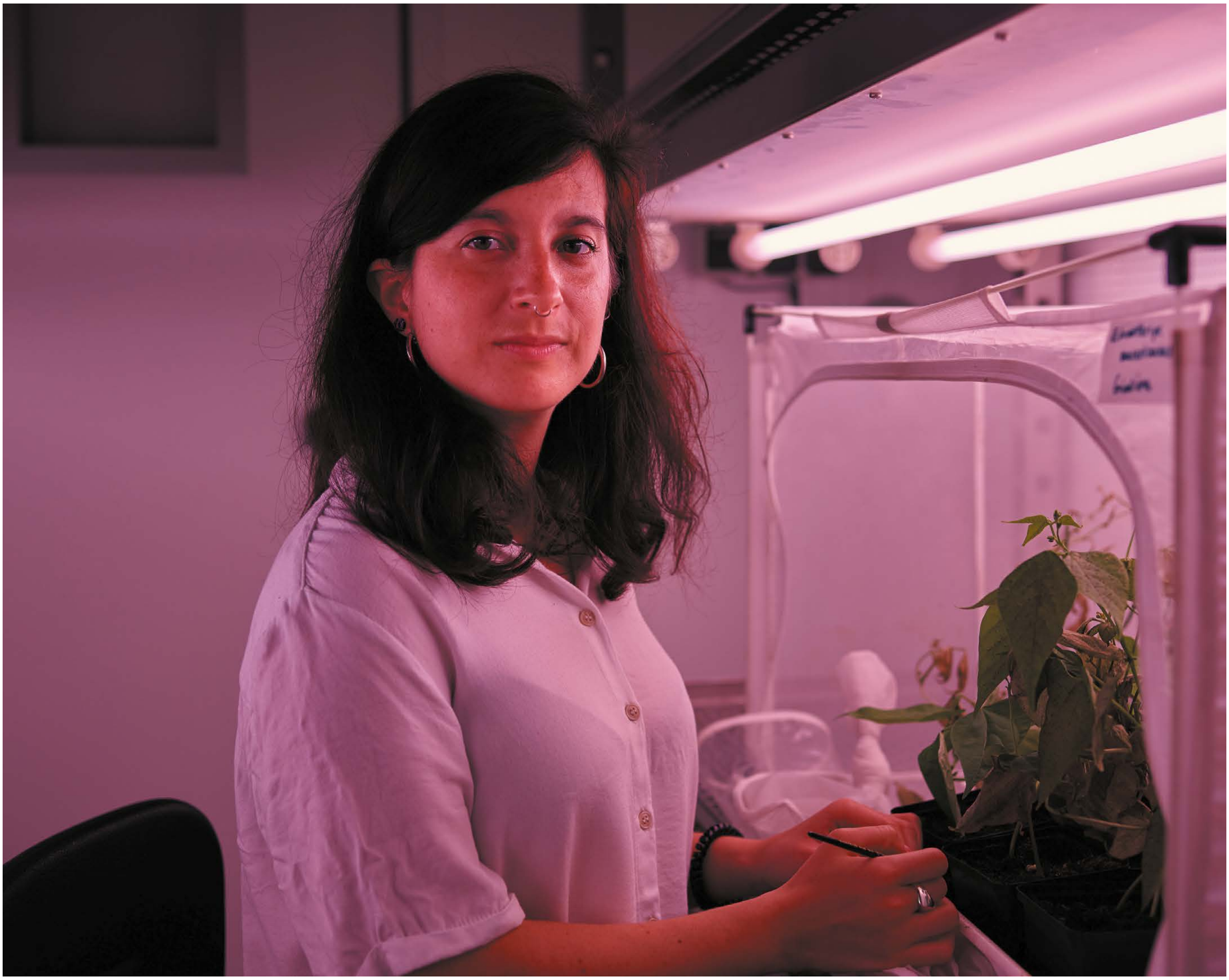
Correction

The brews and bakes that forged career paths

This article misstated the date on which J. Nikol Jackson-Beckham left Randolph College; it was mid-2020.

See https://doi.org/10.1038/d41586-020-02404-3

The back page



Where I work
Giuditta Beretta

Photographed for *Nature* by Niels Jansen/Phenster.

Even in a pandemic, my thrips come first. I worried more about these insects – common garden pests that feast on tulips, roses and other important crops here in the Netherlands – than I worried about myself. Here, I'm using a brush to gently herd hundreds of thrips (*Echinothrips americanus*) onto their new home, a bean plant. Later, I'll try to kill them with predatory mites, a potential biological weapon that could be deployed in greenhouses. For now, I just want them to be healthy.

This climate-controlled chamber at the University of Amsterdam is a thrip paradise. It's a constant 25°C with 75% humidity. The purple light helps the bean plant to grow. A plain white light would probably suffice, but we decided we should do something nice for the plant after covering it in pests.

When the university partly shut down for nearly two months during the pandemic, starting in mid-March, I couldn't do any of my mite experiments, but I was allowed to visit the thrips once a week. They thrived. They're pretty hard to kill, as many gardeners know.

The predatory mites are trickier to keep alive. Instead of leaving them to fend for themselves for a week at a time, I took a bunch home in a plastic container. I fed them a mixture of even smaller mites and yeast, which they like.

We're doing experiments with two types of thrip predator: plant mites (*Amblyseius swirskii*) and various species of soil mite. Plant mites are tiny, and it's comical to see them try to wrap their legs around a thrip in an attack. The soil mites are about 2 millimetres long, nearly the size of the thrips themselves. We hope that they might be more effective against thrips, but it's hard to get soil mites to climb up a plant. A combination might prove most effective.

Although I spend time caring for my thrips, I have no problem killing them. They are not nice insects, and they look ugly under the microscope. I used to work with caterpillars, and I definitely felt more guilty about them.

Giuditta Beretta is a PhD student in evolutionary and population biology at the University of Amsterdam in the Netherlands. **Interview by Chris Woolston.**

APP gene copy number changes reflect exogenous contamination

https://doi.org/10.1038/s41586-020-2522-3

Received: 16 July 2019

Accepted: 18 May 2020

Junho Kim[1,2,3], Boxun Zhao[1,2,3], August Yue Huang[1,2,3], Michael B. Miller[1,2,3,4,5,6], Michael A. Lodato[1,2,3,4,5,7], Christopher A. Walsh[1,2,3,4,5] ✉ & Eunjung Alice Lee[1,2,3] ✉

ARISING FROM M. H. Lee et al. *Nature* https://doi.org/10.1038/s41586-018-0718-6 (2018)

Various types of somatic mutations occur in cells of the human body and cause human diseases, including cancer and some neurological disorders[1]. Recently, Lee et al.[2] (hereafter 'the Lee study') reported somatic copy number gains of the *APP* gene, a known risk locus for Alzheimer's disease (AD), in 69% and 25% of neurons of AD patients and controls, respectively, and argued that the mechanism of these copy number gains was somatic integration of *APP* mRNA into the genome, creating what they called genomic cDNA (gencDNA). Our reanalysis of the data from the Lee study and two additional whole-exome sequencing (WES) data sets by the authors of the Lee study[3] and Park et al.[4] revealed evidence that *APP* gencDNA originates mainly from exogenous contamination by *APP* recombinant vectors, nested PCR products, and human and mouse mRNA, respectively, rather than from true somatic integration of endogenous *APP*. We further present our own single-cell whole-genome sequencing (scWGS) data that show no evidence for somatic *APP* retrotransposition in neurons from individuals with AD or from healthy individuals of various ages.

We examined the original *APP*-targeted sequencing data from the Lee study to investigate sequence features of *APP* retrotransposition. These expected features included (a) reads spanning two adjacent *APP* exons without intervening intron sequence, which would indicate processed *APP* mRNA, and (b) clipped reads, which are reads spanning the source *APP* and new genomic insertion sites, thus manifesting partial alignment to both the source and target site (Extended Data Fig. 1a). The first feature is the hallmark of retrogene or pseudogene insertions, and the second is the hallmark of RNA-mediated insertions of all kinds of retroelements, including retrogenes as well as LINE1 elements. We indeed observed multiple reads spanning two adjacent *APP* exons without the intron; however, we could not find any reads spanning the source *APP* and a target insertion site. Unexpectedly, we found multiple clipped reads at both ends of the *APP* coding sequence that contained the multiple cloning site of the pGEM-T Easy Vector (Promega), which indicates external contamination of the sequencing library by a recombinant vector carrying an insert of *APP* coding sequence (Fig. 1a). The *APP* vector we found here was not used in the Lee study, but rather had been used in the same laboratory when first reporting genomic *APP* mosaicism[5], suggesting carryover from the prior study.

Recombinant vectors with inserts of gene coding sequences (typically without introns or untranslated regions (UTRs)) are widely used for functional gene studies. Recombinant vector contamination in next-generation sequencing is a known source of artefacts in somatic variant calling, as sequence reads from the vector insert confound those from the endogenous gene in the sample DNA[6]. We have identified multiple incidences of vector contamination in next-generation sequencing data sets from different groups, including our own laboratory (Extended Data Fig. 1b), demonstrating the risk of exposure to vector contamination. In an unrelated study on somatic copy number variation in the mouse brain[7], from the same laboratory that authored the Lee study, we found contamination by the same human *APP* pGEM-T Easy Vector in mouse single-neuron WGS data (Extended Data Fig. 1c). We also observed another vector backbone sequence (pTriplEx2, SMART cDNA Library Construction Kit, Clontech) with an *APP* insert (Extended Data Fig. 1c, magnified panel) in the same mouse genome data set, indicating repeated contamination by multiple types of recombinant vectors in the laboratory.

PCR-based experiments with primers that target the *APP* coding sequence (for example, Sanger sequencing and SMRT sequencing) are unable to distinguish *APP* retrocopies from vector inserts (Fig. 1a, top). Therefore, to definitively distinguish between the three potential sources of *APP* sequencing reads (original source *APP*, retrogene copy, and vector insert), it is necessary to study non-PCR-based sequencing data (for example, SureSelect hybrid-capture sequencing) and to examine reads at both ends of the *APP* coding sequence. Such data can help to clarify whether the clipped sequences map to a new insertion site or to vector backbone sequence (Fig. 1a, bottom). From the SureSelect hybrid-capture sequencing data in the Lee study, we directly measured the level of vector contamination by calculating the fraction of the total read depth at both ends of the *APP* coding sequence that consisted of clipped reads containing vector backbone sequences (Fig. 1b, red dots). Similarly, we measured the clipped read fraction at each *APP* exon junction, which indicates the total amount of *APP* gencDNA (either from *APP* retrocopies or vector inserts) (Fig. 1b, black dots). The average clipped read fraction at coding sequence ends that contained vector backbones (1.2%, red dots) was comparable to the average clipped read fraction at exon junctions (1.3%, black dots; $P = 0.64$, Mann–Whitney *U* test), suggesting that vector contamination was the primary source of the clipped reads across all the exon junctions. Even including these vector-originating reads, all the fractions at every junction are far below the conservative estimate of 16.5% gencDNA contribution based on the Lee study's DNA in situ hybridization (DISH) experimental results, which are from the same samples (see Supplementary Information for more details on the discrepancy between sequencing and DISH results). It is incumbent on the authors to provide an explanation for this inconsistency. Moreover, if the clipped reads were from endogenous retrocopies, the clipped and non-clipped reads would be expected to have a similar insert (DNA fragment) size distribution; however, in the Lee study, the clipped reads had a significantly smaller and far more homogeneous insert size distribution than the non-clipped reads that

---

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, USA. [2]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA. [5]Department of Neurology, Harvard Medical School, Boston, MA, USA. [6]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [7]Present address: Department of Molecular, Cell, and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA. ✉e-mail: christopher.walsh@childrens.harvard.edu; ealice.lee@childrens.harvard.edu

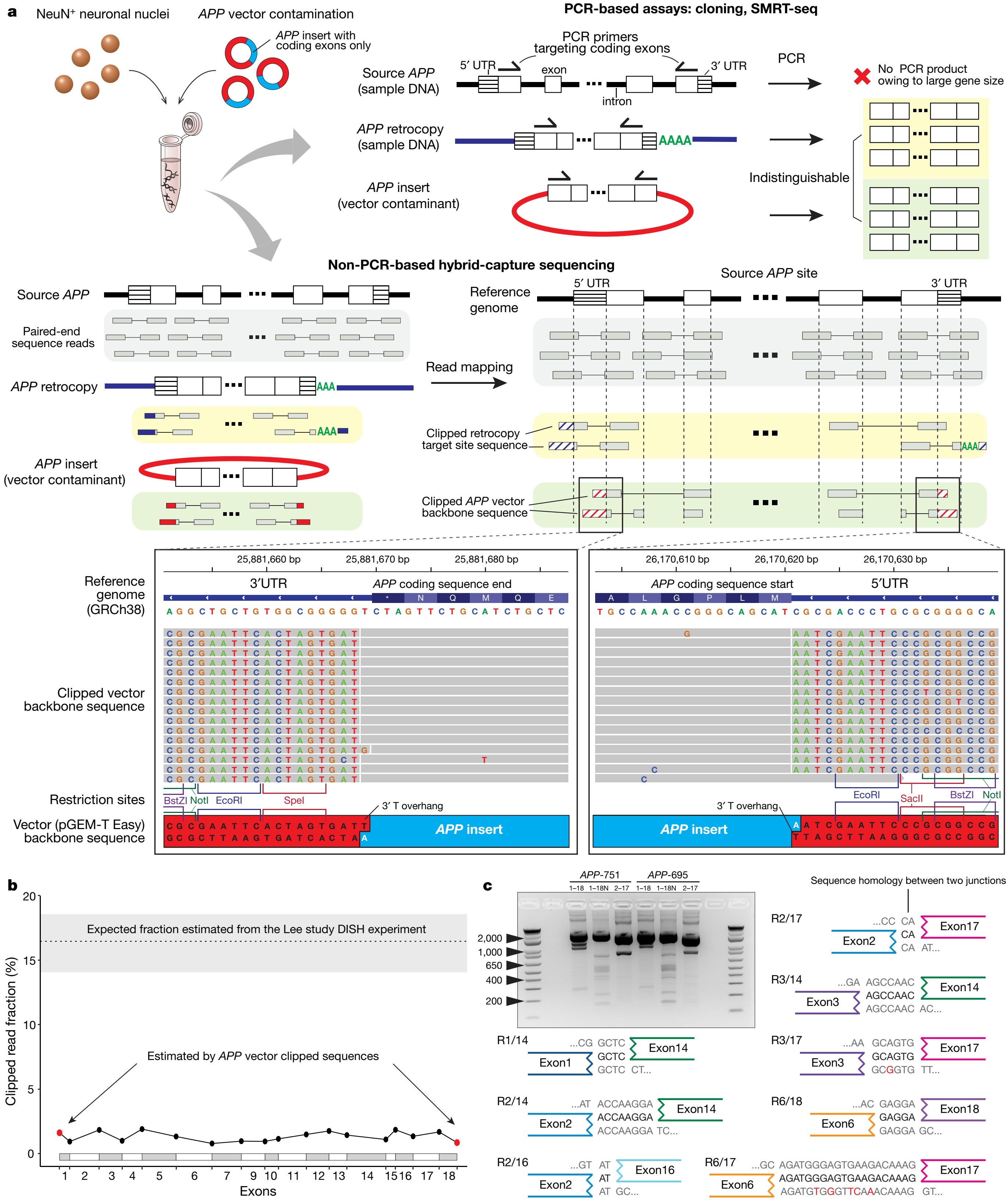


Fig. 1 | *APP* vector contamination in the Lee study. a, *APP* vector contamination and its manifestation in genome sequences. PCR-based assays in the Lee study[2] fail to distinguish between *APP* retrocopy and vector *APP* insert. Hybrid-capture sequences from the Lee study show clipped reads with a vector backbone sequence (pGEM-T Easy), including restriction sites at the multiple cloning site and a 3′ T-overhang. **b**, Estimated fractions of cells with *APP* gencDNA at the exon junctions in the Lee hybrid-capture data. All exon junction fractions (black dots) are comparable to the fraction at the coding sequence ends with vector backbone sequences (red dots). The dotted line above represents the conservative estimate of expected fraction based on the Lee DISH experiment (see Supplementary Methods); shaded area, 95% confidence interval. **c**, Electrophoresis and sequencing of PCR products from the vector *APP* inserts (*APP*-751/695) showing new *APP* variants as artefacts. Eight out of twelve IEJs found both in our *APP* vector PCR sequencing and the Lee study RT–PCR results are shown (Extended Data Fig. 3).

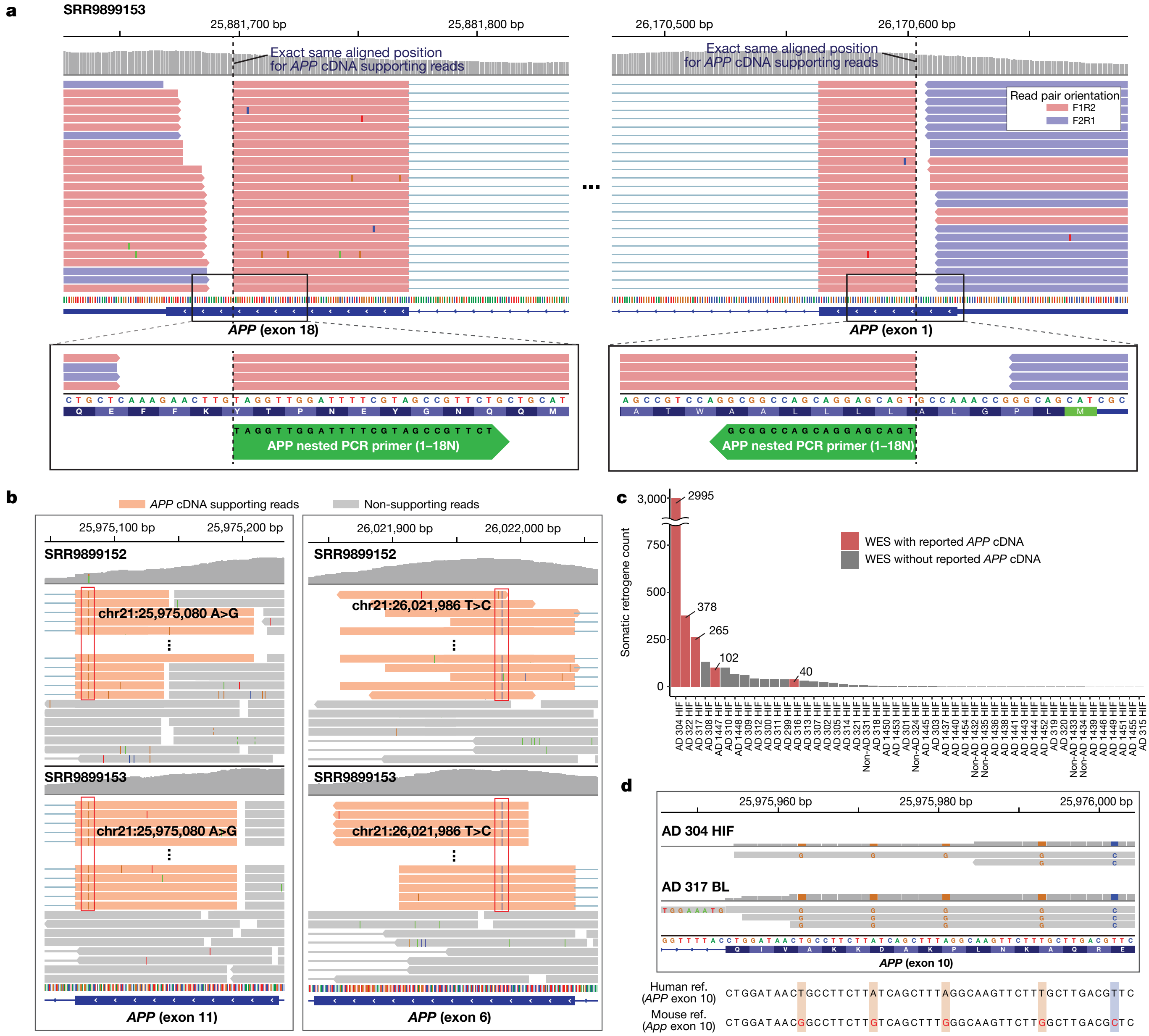


Fig. 2 | *APP* cDNA-supporting reads originate from exogenous PCR products and genome-wide human and mouse mRNA contamination. **a**, *APP* nested PCR products found in the recent Lee WES data[3]. Reads that support *APP* cDNA are aligned to the target sites (dotted lines) of the nested PCR primers (green arrows at the bottom) used in the original Lee study[2]. All these cDNA-supporting reads contain an IEJ between exons 2 and 17 (full structure not shown). **b**, The same unannotated variants found at two different positions (red boxes) only in cDNA-supporting reads (orange) in both WES data sets by Lee et al. (SRR989152 and SRR989153)[2,3]. **c**, Total gene counts with potential somatic retrogene insertions in the Park et al. data[4]. WES data with reported *APP* cDNA are marked in red. **d**, *APP* cDNA-supporting reads originating from mouse mRNA in the Park data. Mouse-specific single-nucleotide polymorphisms (coloured bases) are observed in a portion of cDNA-supporting reads, including those with clipped sequences for exon–exon junctions, suggesting the reads originated from mouse mRNA rather than genomic DNA (Supplementary Fig. 1).

were from original source *APP*, thus demonstrating the foreign nature of the clipped reads ($P < 2.2 \times 10^{-16}$, Mann–Whitney *U* test; Extended Data Fig. 2a–c, see Supplementary Information). Finally, we found no direct evidence to support the existence of true *APP* retrogene insertions, such as clipped and discordant reads near the *APP* UTR ends that mapped to a new insertion site, or clipped reads with polyA tails at the 3′ end of the UTR, although the sequencing depth of UTRs was over 500×. Given that the hybrid capture experiment appears properly designed to detect *APP* gencDNA, the absence of any bona fide insertion signal suggests the absence of true *APP* gencDNA and that the majority of *APP*-gencDNA-supporting reads originated from *APP* vector contamination.

The authors of the Lee study have subsequently generated WES data sets from the brain samples of six patients with AD and one control individual without AD (Sequence Read Archive (SRA) accession: PRJNA558504), and reported multiple reads spanning *APP* exons without introns as evidence of somatic *APP* gencDNA[3]. We confirmed this in the data, but again, found not a single read spanning the source *APP* and any insertion sites. Instead, the data revealed anomalous patterns in a subset of reads supporting *APP* gencDNA. Those reads spanning exons 1 and 18 were aligned to the exact same start and end positions with the same read pair orientation (Fig. 2a), which is unlikely to occur in non-PCR-based exome capture sequencing. We found that the two aligned positions within exons 1 and 18 exactly matched the target

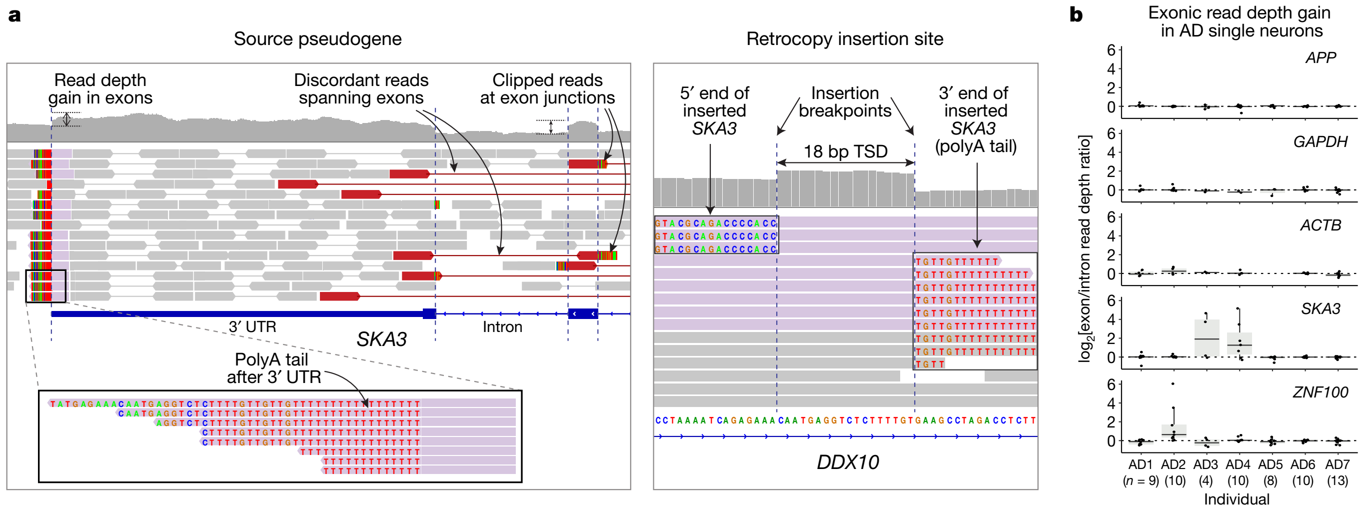


Fig. 3 | Absence of somatic *APP* retrogene insertions in our scWGS data. **a**, A germline pseudogene insertion (*SKA3*) in our scWGS data showing all distinctive characteristics of true retrogene insertion. **b**, No read-depth gain in *APP* exons in our single neurons from patients with AD. Each dot represents the median of exon/intron read-depth ratios across all exons of the gene in each scWGS data set from patients with AD. Patients with AD who have polymorphic germline retrogene insertions of *SKA3* (AD3 and AD4) or a germline insertion of *ZNF100* (AD2) show clear read-depth gain; there is no such gain for two housekeeping genes (*GAPDH*, *ACTB*). Single cells that had poor genomic coverage for a given gene due to locus dropout are excluded. *n*, number of single cells in each individual; centre line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range.

sites of the nested PCR primers used in the original Lee study (1-18N, Supplementary Table 1 in the Lee study). The only explanation for this observation is contamination of the WES library by nested PCR products from the original APP study. This finding raises serious concerns that *APP* PCR products may also have contaminated the genomic DNA samples and were fragmented and sequenced together, generating more gencDNA-compatible reads for which we are unable to clarify the source. We also identified two unannotated (that is, absent in the gnomAD) single-nucleotide variants in all *APP*-cDNA-supporting reads in the two independent WES libraries pooled from six AD samples, which is very unlikely to be observed in different individuals, thus supporting the possibility that the *APP* cDNA originated from the same external source (Fig. 2b).

An independent study by Park et al.[4] has recently presented a small fraction of reads supporting *APP* cDNA in deep WES data sets from AD brain samples (SRA accession: PRJNA532465; Supplementary Fig. 12 in the study). These data were free from vector contamination, but we found evidence of genome-wide human mRNA contamination, predominantly in the WES data sets with reported *APP* cDNA supporting reads. We note that their analysis of somatic single-nucleotide variants (SNVs) is likely to be unaffected by this contamination owing to their visual inspection and stringent filtering of known germline SNVs. For each AD brain sample, we counted the number of genes with potential somatic retrotransposition events by checking whether a gene had cDNA-supporting reads (that is, reads connecting two adjacent exons and skipping the intervening intron) at more than two different exon junctions in the brain sample but not in the matched blood sample from the same patient (see Supplementary Methods). All WES data sets reported by the authors to have *APP* cDNA showed an extremely high number of other genes in addition to *APP* with cDNA-supporting reads (40–2,995 genes; Fig. 2c). Considering that far fewer than one somatic retrogene insertion per sample would be expected for human cells, even for human cancers with a high rate of somatic LINE1 retrotransposition (for example, lung and colorectal cancer)[8], this result strongly suggests that cDNA-supporting reads could not have originated from true somatic insertions of hundreds to thousands of retrogenes but rather supports the presence of genome-wide human mRNA contamination. We also found cDNA-supporting reads, including a subset of *APP* cDNA-supporting reads, that originated from mouse mRNA, additionally confirming mRNA contamination of the data (Fig. 2d, Supplementary Fig. 1). We observed mRNA contamination in one cell in our scWGS data (see Supplementary Information). Neither Park et al. (personal communication) nor we had performed any mRNA experiments, suggesting that contamination might have arisen from a source outside the research laboratories, such as the sequencing facility. We found no evidence of genuine *APP* genomic cDNA either in the new WES data from the Lee study authors, or in the independent Park et al. data. These findings highlight pervasive exogenous contamination in next-generation sequencing experiments, even with high quality-control standards, and emphasizes the need for rigorous data analysis to mitigate these important sources of artefacts.

The Lee study reported numerous new forms of *APP* splice variants with intra-exon junctions (IEJs), with greater diversity in patients with AD than in healthy individuals. The authors also presented short sequence homology (2–20 bp) at IEJs and suggested that microhomology-mediated end-joining contributed to IEJ formation. It is well known that microhomology can predispose to PCR artefacts[9], and the Lee study performed a high number of PCR cycles in their experimental protocol (40 cycles). Thus, we tested the hypothesis that the IEJs in the Lee study could have arisen as PCR artefacts from the PCR amplification of a contaminant. To do so, we repeated in our laboratory both RT–PCR and PCR assays following the Lee study protocol using recombinant vectors with two different *APP* isoforms (*APP*-751, *APP*-695), and using the reported PCR primer sets with three different PCR enzymes as described in their study (see Supplementary Information). Indeed, with all combinations of *APP* inserts and PCR enzymes, we observed chimeric amplification bands with various sizes that were clearly distinct from the original *APP* inserts (Fig. 1c, Extended Data Fig. 3a). We further sequenced these non-specific amplicons and confirmed that they contained numerous IEJs of *APP* inserts (Supplementary Table 1). Twelve of seventeen previously reported IEJs in the Lee study were also found from our sequencing of PCR artefacts (Fig. 1c, Extended Data Fig. 3b). Our observations suggest that the new *APP* variants with IEJs from the Lee study might have originated from contaminants as PCR artefacts. This possibility is corroborated by the fact that IEJ-supporting reads were completely absent from the hybrid-capture sequencing data from the Lee study, and that reads supporting an IEJ in the new WES data set by the authors originated from external nested *APP* PCR products (Fig. 2a).

To independently investigate potential *APP* gencDNA, we searched for somatic *APP* retrogene insertions in our independent scWGS data from patients with AD and healthy control individuals. In brief, we isolated single neuronal nuclei using NeuN staining followed by fluorescence-activated cell sorting (FACS), amplified the whole genome using multiple displacement amplification (MDA), and finally sequenced the whole genome at 45× mean depth[10]. The dataset consists of a total of 64 scWGS data sets from 7 patients with Braak stage V and VI AD, along with 119 scWGS data sets from 15 unaffected control individuals, some of which have been previously published[11]. Our previous studies and those by other groups[10,12–14] have successfully detected and fully validated bona fide somatic insertions of LINE1 by capturing distinct sequence features in scWGS data, demonstrating the high resolution and accuracy of scWGS-based retrotransposition detection. Therefore, if a retrogene insertion had occurred, we should have been able to observe distinct sequence features at the source retrogene site: increased exonic read-depth, read clipping at exon junctions, poly-A tail at the end of the 3′ UTR, and discordant read pairs spanning exons (Extended Data Fig. 1a). We captured these features at the existing germline retrogene insertions, such as the *SKA3* pseudogene insertion (Fig. 3a). If present, somatic events should be able to be detected as heterozygous germline variants in scWGS; however, our analysis revealed no evidence of somatic *APP* retrogene insertions in any cell. By contrast, in both patients (AD3 and AD4) with germline insertions of *SKA3* and the patient (AD2) with a germline insertion of *ZNF100*, there was a clear increase in exonic read depth relative to introns, as would signal for polymorphic germline retrogene insertions (Fig. 3b). We observed no such read depth increase for *APP* in our 64 AD and 119 normal single-neuron WGS profiles, confirming that we found no evidence of *APP* retrogene insertions in human neurons.

In summary, our analysis of the original sequencing data from the Lee study, the new WES data from the same authors, and the WES data from the independent Park study, as well as of our own scWGS data, suggests that somatic *APP* retrotransposition does not frequently occur in neurons from either patients with AD or healthy individuals. Rather, the reported evidence of *APP* retrocopies appears to be attributable to various types of exogenous contamination—specifically *APP* recombinant vectors, PCR products, and genome-wide mRNA contamination. Our replication experiment also showed that it is possible for PCR amplification artefacts to create spurious products that mimic *APP* gene recombination with various internal exon junctions. Thus, to support the claimed phenomenon of *APP* gencDNA, it would be necessary for the authors to present unequivocal evidence that cannot be attributed to contamination, such as reads that support new *APP* insertion breakpoints; however, the authors have not presented such direct evidence. In conclusion, we found no evidence of *APP* retrotransposition in the genomic data presented in the Lee study and further show that our own single-neuron WGS analysis, which directly queried the *APP* locus at single-nucleotide resolution, reveals no evidence of *APP* retrotransposition or insertion.

Data availability

APP vector PCR sequences have been deposited in the NCBI SRA (PRJNA577966). Single-cell whole-genome sequencing data from control individuals have been deposited in the NCBI SRA (PRJNA245456) and dbGAP (phs001485.v1.p1). Single-cell whole-genome sequencing data from patients with AD are available upon request for genomic regions of *APP* and source pseudogene *SKA3* and *ZNF100*.

Code availability

Implemented custom code for the estimation of clipped read fractions and the detection of intra-exon junctions (IEJs) is available at https://sourceforge.net/projects/somatic-app-analysis/.

1. McConnell, M. J. et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, eaal1641 (2017).
2. Lee, M. H. et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639–645 (2018).
3. Lee, M.-H. et al. Reply: *APP* gene copy number changes reflect exogenous contamination. *Nature* https://doi.org/10.1038/s41586-020-2523-2 (2020).
4. Park, J. S. et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun*. **10**, 3090 (2019).
5. Bushman, D. M. et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife* **4**, (2015).
6. Kim, J. et al. Vecuum: identification and filtration of false somatic variants caused by recombinant vector contamination. *Bioinformatics* **32**, 3072–3080 (2016).
7. Rohrback, S. et al. Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **115**, 10804–10809 (2018).
8. Cooke, S. L. et al. Processed pseudogenes acquired somatically during cancer development. *Nat. Commun*. **5**, 3644 (2014).
9. Odelberg, S. J., Weiss, R. B., Hata, A. & White, R. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Res*. **23**, 2049–2057 (1995).
10. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
11. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
12. Erwin, J. A. et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci*. **19**, 1583–1591 (2016).
13. Evrony, G. D., Lee, E., Park, P. J. & Walsh, C. A. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5**, e12966 (2016).
14. Zhao, B. et al. Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLoS Genet*. **15**, e1008043 (2019).
15. Zhang, X. et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* **166**, 1147–1162.e1115 (2016).

Acknowledgements E.A.L. is supported by grants from the NIA (K01AG051791), the Suh Kyungbae Foundation, and the Charles H. Hood foundation. This work was also supported by the Paul G. Allen Frontiers Group (C.A.W., E.A.L.), NINDS grant R01NS032457-20S1 (C.A.W.), DOD grant W18XWH2010028 (J.K., E.A.L., C.A.W.), Manton Center Pilot Project Award and Rare Disease Research Fellowship (B.Z.), NIH grants T32HL007627 and K08AG065502 (M.B.M.), and NIH grant AG054748 (M.A.L). C.A.W. is an Investigator of the Howard Hughes Medical Institute.

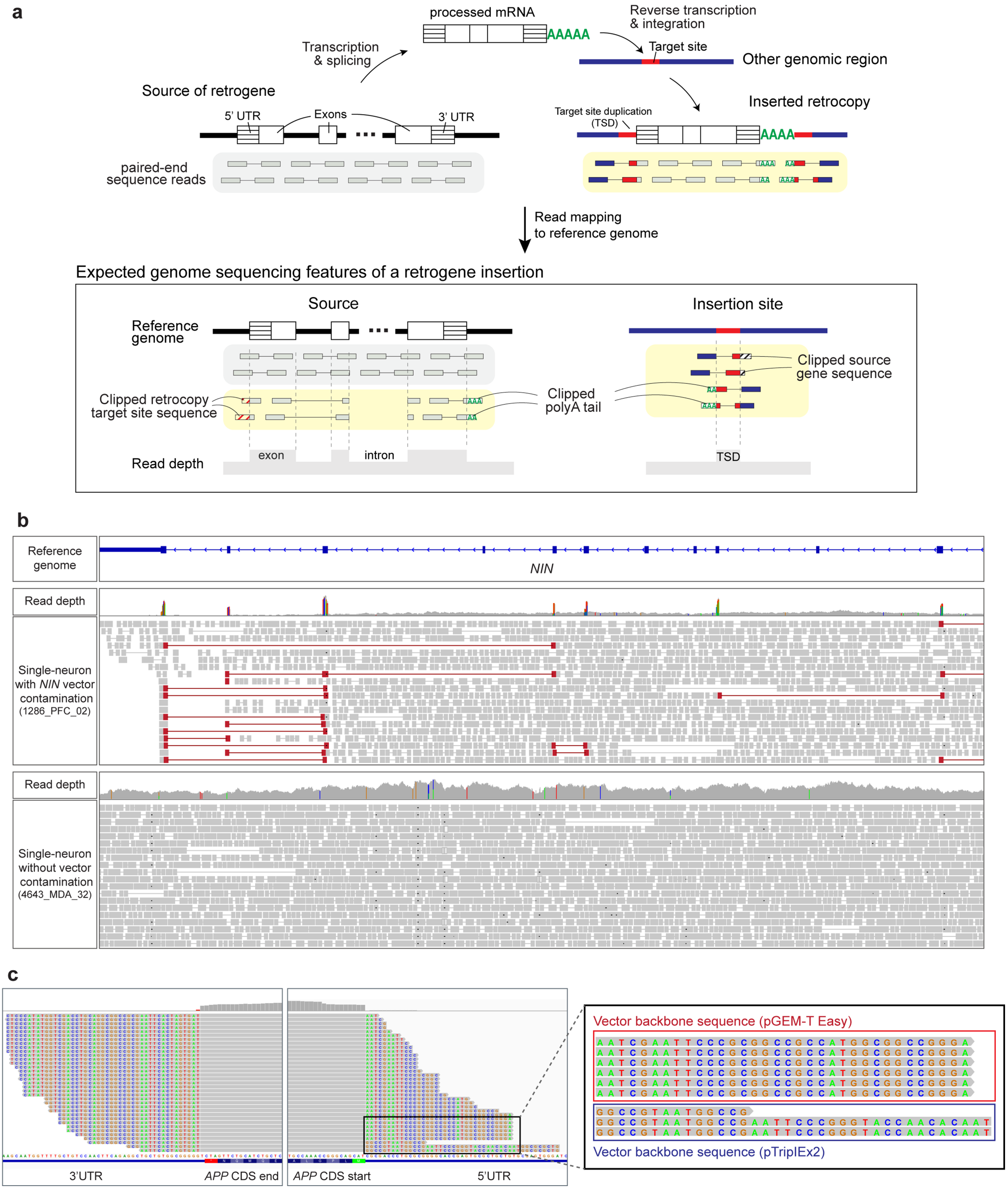
Author contributions J.K. and E.A.L. conceived and designed the study. J.K. and B.Z. designed the *APP* vector PCR and sequencing, and B.Z. performed the PCR and sequencing. M.B.M. and M.A.L. performed single-neuron sorting and sequencing. J.K. and A.Y.H. performed bioinformatic analyses. E.A.L and C.A.W supervised the study. J.K., B.Z., M.B.M., M.A.L., C.A.W., and E.A.L. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

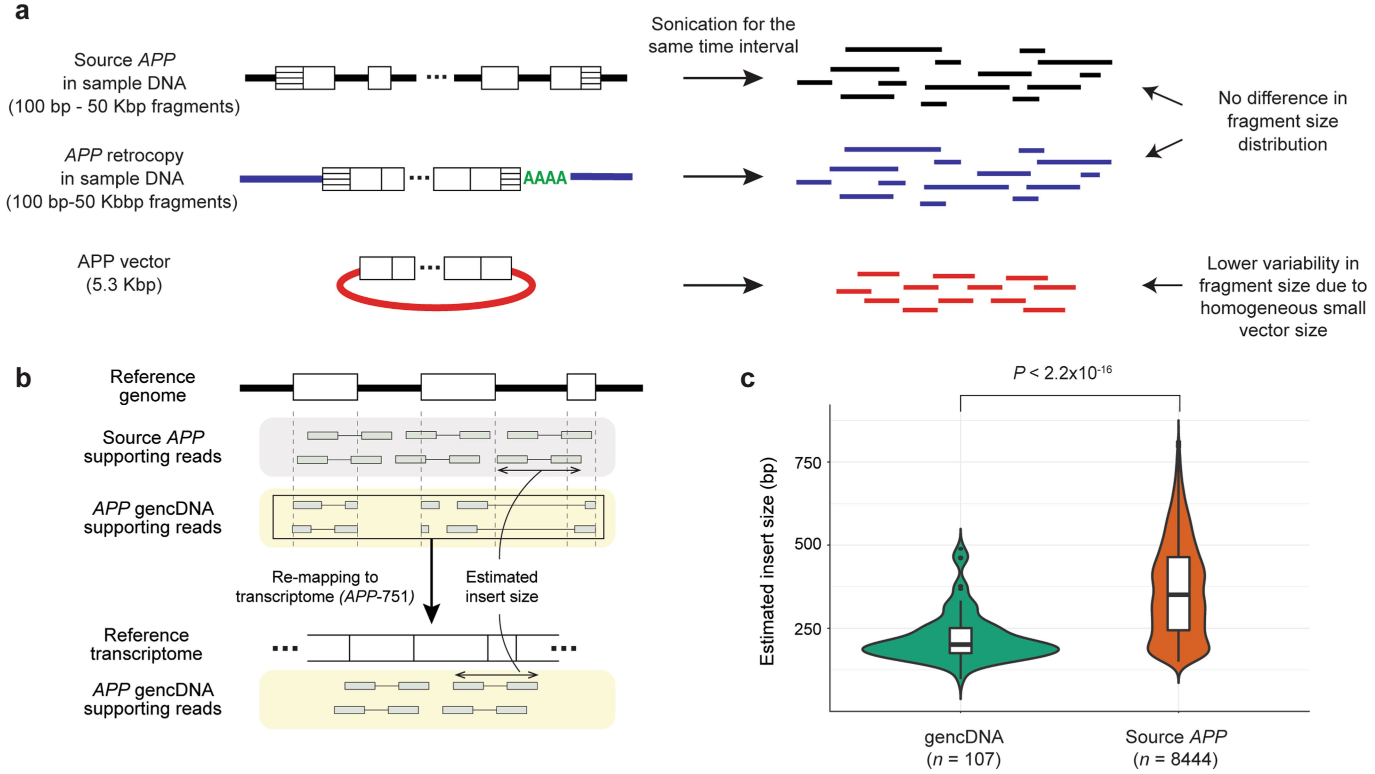
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2522-3.

Correspondence and requests for materials should be addressed to C.A.W. or E.A.L.

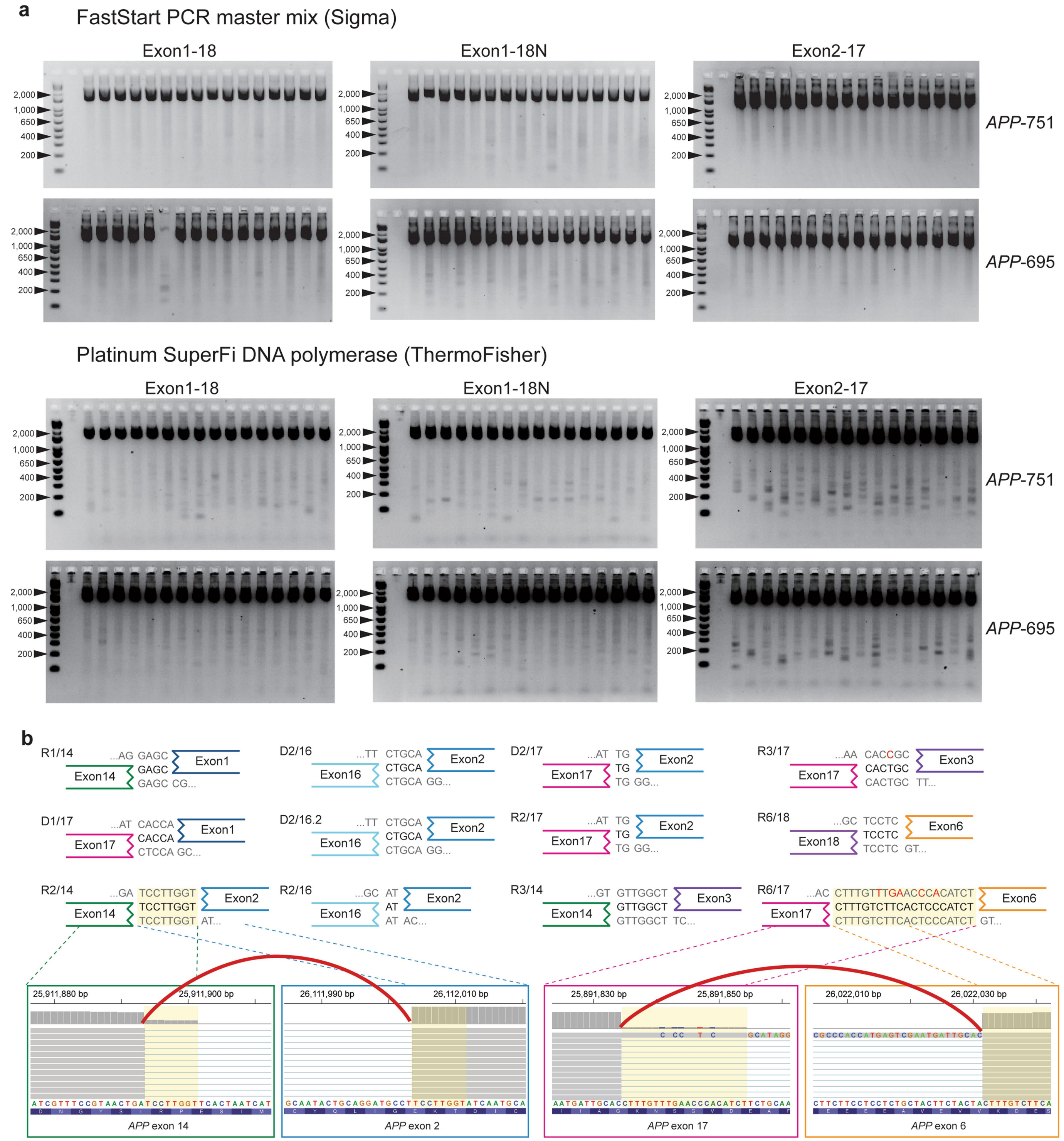


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Pervasive recombinant vector contamination in next-generation sequencing. a, Schematic of a retrogene insertion and the characteristics expected to be captured in sequencing data: increased exonic read-depth, discordant reads spanning exons, clipped reads at exon junctions, 3′ poly-A tail, target site duplication (TSD) at the new genomic insertion site, and clipped reads spanning the retrocopy and insertion sites. **b**, Recombinant vector contamination found in the Walsh laboratory data. Four single human neurons (1286_PFC_02, 1762_PFC_04, 5379_PFC_01, 5416_PFC_06) in our previous publication showed contamination by a mouse *Nin* recombinant vector[15]. The homologous human gene region (*NIN*) is visualized by the IGV browser for a vector-contaminated cell (top) and an unaffected control cell (bottom). Contamination characteristics were identified, including increased exonic read-depth and exon-spanning discordant reads (reads coloured in red) with numerous mismatches to the human genome reference (coloured vertical bars in the read depth track). **c**, Mouse single-neuron WGS data from the Chun laboratory[7] contaminated by the same *APP* recombinant vector detected in the Lee study[2] and an additional *APP* plasmid vector (magnified panel).



Extended Data Fig. 2 | Evidence that recombinant vector contamination is the major source of *APP* gencDNA. a, Schematic of the DNA fragment size distribution for each *APP* source (source *APP*, *APP* retrocopy, *APP* vector). Fragments from *APP* vectors are expected to be more homogeneous and smaller than those from other sources owing to the fixed and relatively small vector size. **b**, DNA fragment (or insert) size estimation. Sequence reads mapped to *APP* exon junctions were divided into two groups: source *APP* (reads containing intron sequences) and *APP* gencDNA (reads clipped at the exon junction) supporting reads. gencDNA supporting reads were remapped to the *APP* reference transcript sequence (*APP*-751) to estimate insert sizes. **c**, Comparison of insert size distribution between source and gencDNA supporting reads. *n*, number of read pairs in each group; centre line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range.



Extended Data Fig. 3 | New *APP* variants with intra-exon junctions as PCR artefacts. a, Electrophoresis of PCR products from the vector *APP* inserts (*APP*-751, *APP*-695) showing novel *APP* variants as artefacts. All combinations of two PCR enzymes (FastStart PCR master mix and Platinum SuperFi DNA polymerase; OneStep Ahead RT–PCR in Fig. 1c) with three primer sets generated new bands smaller than the expected PCR product. **b**, PCR-induced IEJs with homologous sequences at each junction identified by Illumina sequencing. Twelve IEJs from our vector PCR sequencing showed exactly the same sequence homologies and genomic coordinates as IEJs reported by Lee et al[2]. For two IEJs, IGV browser images show pre- (left) and post-junction sites (right) connected by split reads spanning the IEJ (red arc). Because IGV displays forward strand sequences of the human reference genome, all IEJ sequences were also reverse complemented for consistent visualization.

nature research

Corresponding author(s): Eunjung Alice Lee

Last updated by author(s): Apr 27, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SRA toolkit (2.9.0) was used to download the sequencing data from the Chun laboratory (SRP162675, SRP121019) from the Sequence Read Archive as described in the Supplementary Information.
Data analysis	Sequencing data was processed to generate analysis-ready BAM using Cutadapt (1.1.4), BWA-mem (0.7.17), Picard (2.8.0), and GATK (3.5) as described in the Supplementary Information. Vecuum (1.0.1) and NCBI BLASTN were used to clarify APP vector contamination. Implemented custom code for the calculation of clipped read fractions and the detection of intra-exon junctions will be uploaded to open source repository (SourceForge).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

APP vector PCR sequences have been deposited in the NCBI Sequence Read Archive (PRJNA577966). Single-cell whole genome sequencing data of control individuals have been deposited in the NCBI Sequence Read Archive (PRJNA245456) and dbGAP (phs001485.v1.p1). Single-cell whole genome sequencing data of AD patients will be available upon request for the genomic regions of APP and source pseudogene SKA3 and ZNF100.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We analyzed our independent single-cell whole-genome sequencing (scWGS) data of AD and control neurons including previously published data sets (Lodato et al, Science, 2017). The sample size was determined by the number of sequenced cells (64 scWGS from 7 AD patients and 119 scWGS from 15 unaffected controls). This was sufficient to verify the absence of somatic APP retrotransposition, which was reported as occurring in 69% of AD neurons on average (Binomial P < 2.2e-16).
Data exclusions	One single cell (5087_MDA_02) from the public sequencing data (Lodato et al, Science, 2018) was excluded due to genome-wide mRNA contamination.
Replication	Somatic APP retrotransposition was examined in independent scWGS data from AD patients and normal controls. Both original sequencing data from the Lee study (Lee et al., Nature, 2018) and independent scWGS data show no evidence of somatic APP retrotransposition.
Randomization	Not relevant to our study since we utilized all available data sets without any allocation of samples.
Blinding	Not relevant to our study as no difference between AD and control groups was observed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Matters arising

Reply to: *APP* gene copy number changes reflect exogenous contamination

https://doi.org/10.1038/s41586-020-2523-2

Received: 24 April 2020

Accepted: 18 May 2020

Ming-Hsiang Lee[1,3], Christine S. Liu[1,2,3], Yunjiao Zhu[1], Gwendolyn E. Kaeser[1], Richard Rivera[1], William J. Romanow[1], Yasuyuki Kihara[1] & Jerold Chun[1]✉

REPLYING TO J. Kim et al. *Nature* https://doi.org/10.1038/s41586-020-2522-3 (2020)

In the accompanying comment[1], Kim et al. conclude that somatic gene recombination (SGR) and amyloid precursor protein (*APP*) genomic complementary DNAs (gencDNAs) in the brain are contamination artefacts and do not naturally exist. We disagree. Here we address the three types of analyses used by Kim et al. to reach their conclusions: informatic contaminant identification, plasmid PCR, and single-cell sequencing. Additionally, Kim et al. requested "reads supporting novel *APP* insertion breakpoints," and we now provide ten different examples that support *APP* gencDNA insertion within eight chromosomes beyond wild-type *APP* on chromosome 21 from patients with Alzheimer's disease. If SGR exists, as experimentally supported here and previously[2,3], contamination scenarios become moot.

Our informatic analyses of data generated by an independent laboratory (Park et al.)[4] complement, and are entirely consistent with, what Lee et al.[2] presented via nine distinct lines of evidence, in addition to three from a prior publication[3]. Plasmid contamination was identified in a single pull-down dataset after publication of Lee et al.[2]; however, subsequent analyses did not alter any of our conclusions, including those of our prior publications[3,5], and plasmid contamination-free replication of this approach by ourselves and others supported the original conclusions. Novel retro-insertion sites, alterations of *APP* gencDNA number and form within cell types from the same brain, and pathogenic SNVs that occur only in samples from patients with AD, all support the existence of *APP* gencDNAs produced by SGR.

One predicted outcome of SGR is the generation of novel retro-insertion sites distinct from the wild-type locus, as we demonstrated using DNA in situ hybridization (DISH; Fig. 2n in Lee et al.). Analyses of independently published data sets[4] produced by whole-exome pull-down of DNA from laser-captured human hippocampus or blood revealed ten different *APP* insertion sites within eight different chromosomes (Fig. 1, Supplementary Table 1). We identified clipped reads spanning *APP* untranslated regions (UTRs) and new genomic insertion sites on chromosomes 1, 3, 9, 10, and 12 (Fig. 1a; wild-type *APP* is located on chromosome 21). The corresponding paired-end reads mapped to the same inserted chromosome. We also identified reads spanning *APP* exon–exon junctions of gencDNAs that had mate-reads mapping to other genomic sites on chromosomes 1, 3, 5, 6, and 13 (Fig. 1b). We are unaware of contamination sources that could produce these results that are entirely consistent with our DISH data showing *APP* gencDNA locations distinct from wild-type *APP*. These new *APP* gencDNA insertion sites strongly support the natural occurrence of *APP* gencDNAs.

An *APP* plasmid contaminant (pGEM-T Easy *APP*) was found in our single pull-down dataset; however, we could not definitively determine which *APP* exon–exon reads resulted from gencDNAs as opposed to plasmid contamination, especially in view of the 11 other distinct and uncontaminated approaches that had independently supported and/or identified *APP* gencDNAs. Three other pull-down datasets from our laboratory were informatically analysed and found to contain *APP* gencDNA reads while being free from *APP* plasmid contamination by both VecScreen[6] and subsequent use of the Vecuum script[7] (Fig. 2a, b). Possible external source contamination noted by Kim et al. in two of three data sets could not definitively account for all *APP* exon–exon junctions.

The recent availability of independently generated datasets derived from patients with AD[4] provided a test for the independent reproducibility of *APP* gencDNA identification. Five brain and two blood samples from individuals with sporadic AD (SAD) contained *APP* gencDNA sequences and were shown to be plasmid-free by Vecuum[7] screening (Fig. 2a–e). In addition to exon–exon junction reads and novel insertion sites, we also identified *APP* UTR sequences paired with reads containing *APP* gencDNA exon–exon junctions (Fig. 2d, e). This may be explained by a key experimental design factor: the pull-down probes used by Park et al. contain sequences corresponding to the 5′ and 3′ UTRs of *APP*.

In addition to *APP* plasmid and amplicon contaminants, Kim et al. invoked genome-wide mouse and human mRNA contamination in the Park et al. data set. We cannot address conditions in the Park et al. laboratory but note that it is completely independent of our own. Kim et al. explain this by implicating the generation of DNA from mRNA, which requires reverse transcriptase activity. The Agilent SureSelect pull-down used by Park et al. and in our experiments do not use reverse transcriptase (Fig. 2a and Supplementary Methods), and we are unaware of any mechanism that would generate DNA from RNA in the absence of reverse transcriptase activity under the conditions used. An alternative explanation is the existence of gencDNAs that affect other genes, as we previously detected in non-*APP* intra-exonic junctions (IEJs) found in commercial cDNA Iso-Seq data sets (Extended Data Fig. 1). Additional validation would be required for new genes, but we note that an average of 450 Mb of extra DNA exists within cortical neurons from individuals with AD[3] that could accommodate new gencDNA sequences. Kim et al. invoked genome-wide mouse mRNA contamination in the Park et al. data set to account for *APP* gencDNAs, but this explanation conflicts with the available data. Mouse-specific single nucleotide polymorphisms (SNPs) in the Park et al. data set cannot account for all *APP* gencDNA-supporting reads: five of seven *APP* exon–exon junction sequences do not contain putative mouse-specific SNPs at the specific region reported by Kim et al. (Fig. 3; Kim et al. Fig. 2d). Most critically, the novel *APP* gencDNA insertion sites identified here cannot be explained by genome-wide mRNA contamination.

---

[1]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA. [2]Biomedical Sciences Program, School of Medicine, University of California San Diego, La Jolla, CA, USA. [3]These authors contributed equally: Ming-Hsiang Lee, Christine S. Liu. ✉e-mail: jchun@sbpdiscovery.org

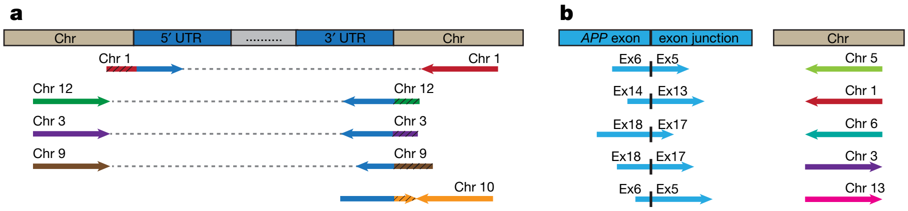


Fig. 1 | Identification of novel *APP* insertion sites in the human genome. **a**, Clipped reads spanning *APP* UTRs and novel chromosomal insertion sites were identified. The paired mate-reads of the clipped reads (black hatching) uniquely mapped to the same chromosomes. **b**, Discordant read-pairs were identified where one read spanned an *APP* exon–exon junction and the corresponding mate-read mapped to a novel chromosome. Each chromosome has a unique colour. Arrowhead direction represents the read orientation after mapping to the human reference genome. Arrows oriented in the same direction support sequence inversions. See detailed sequence and alignment information in Supplementary Table 1.

Kim et al. used PCR of *APP* splice variant plasmids, which generated sequences containing IEJs. However, there are multiple discrepancies between this approach and our biological IEJs and gencDNAs. 1) The experimental conditions, beyond the use of our primer sequences, were different: Kim et al. used twice the concentration of primers and more than one million times more template (250 pg *APP* plasmid is 4.6×10^7 copies versus about 40 gencDNA copies in our PCR of 20 nuclei; based on Lee et al.[2] Fig. 5: DISH 16/17 averaged about 1.8 copies per SAD nucleus). 2) Both gencDNA and IEJ sequences can be detected with as few as 30 cycles of PCR, as we used in single molecule real-time sequencing (SMRT-seq) (Lee et al.[2] Fig. 3) versus 40 cycles used by Kim et al. 3) The agarose gels in Kim et al. are uniformly and unambiguously dominated by a vastly over-amplified about 2-kb band (Kim et al. Fig. 1c and Extended Data Fig. 3a) that is never seen in human neurons despite our routine identification of myriad smaller bands (compare with Lee et al.[2] Fig. 2b). We did observe an over-amplified about 2-kb band in our purposeful plasmid transfection experiments, which also used PCR; however, the formation of gencDNA and IEJs was comparatively limited, of sequences distinct from brain and critically, required both reverse transcriptase activity and DNA strand breakage (Lee et al.[2], Fig. 4). 4) Finally, only 45 unique IEJs from the brains of individuals with AD and 20 from the brains of healthy controls were identified (Lee et al.[2] Fig. 3 with some overlap, fewer than 65 total) compared to the 12,426 identified by Kim et al. (an approximately 200-fold increase over biological IEJs; Kim et al. Supplementary Table 1). We wish to note that microhomology regions within *APP* exons are intrinsic to the *APP* DNA sequence and that microhomology-mediated repair mechanisms involve DNA polymerases[8,9]. The PCR results of Kim et al. differ from our biological data but might inadvertently support the endogenous formation of at least some IEJs within DNA rather than requiring RNA.

Despite these differences between the non-biological plasmid PCR data generated by Kim et al. and our data, Kim et al. conclude that IEJs from our original study[2] might have originated from contaminants. To eliminate this possibility, Lee et al.[2] presented four lines of evidence for *APP* gencDNAs containing IEJs that are independent of *APP* PCR: two different commercially produced cDNA SMRT-seq libraries, DISH, and RNA in situ hybridization (RISH). The SMRT-seq libraries revealed IEJs within *APP* (Lee et al.[2] Extended Data Fig. 1e) as well as other genes (Extended Data Fig. 1), which cannot be attributed to plasmid contamination or PCR amplification. The DISH and RISH results support the existence of *APP* gencDNAs and IEJs (see Supplementary Discussion and Lee et al.[2] Fig. 2, Extended Data Figs. 1, 2) by using custom-designed and validated commercial probe technology (Advanced Cell Diagnostics, ACD), which was independently shown to detect exon–exon junctions[10] and single-nucleotide mutations[11]. Thus, gencDNAs and IEJs can be detected in the absence of targeted PCR. Notably, the contamination proposed by Kim et al. cannot account for the marked change in the number and forms of *APP* gencDNAs that occurs with disease state. The change is also apparent when comparing cell types; signals are vastly

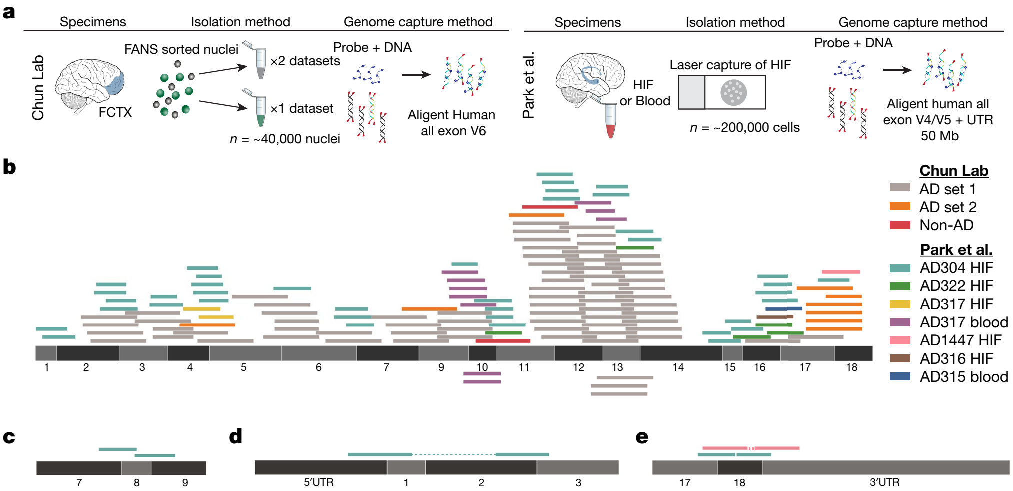


Fig. 2 | Identification of *APP* gencDNA sequences in ten new whole-exome pull-down datasets from two independent laboratories. **a**, Method schematic depicting the standard protocol for whole-exome pull-downs and highlighted methodological differences between the independent laboratories (our lab and Park et al.[4]). **b**, *APP-751* sequence with non-duplicate gencDNA reads from the ten new datasets; colour key indicates the source reads for all panels. **c**, Reads that map to junctions between *APP* exons 7, 8, and 9 that are absent from *APP-751*. **d**, **e**, Paired reads that represent a DNA fragment containing both an exon–exon junction and an *APP* 5′ or 3′ UTR.

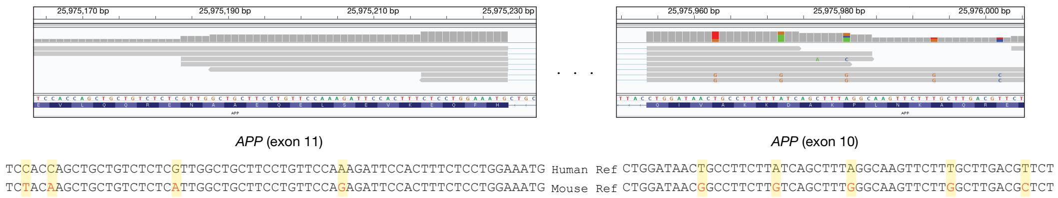


Fig. 3 | Five *APP* gencDNA-supporting reads that span exon–exon junctions and do not contain mouse-specific SNPs. *APP* gencDNA reads were identified that span the *APP* exon10–exon11 junction from the Park et al. datasets[4]. The reference sequences of human and mouse exons are indicated and the positions where the nucleotides differ are highlighted. Five of the seven exon–exon junction-spanning reads do not contain mouse-specific SNPs.

more prevalent in neurons than in non-neuronal cells from the same brains of individuals with SAD when the samples are processed at the same time by DISH (Lee et al.[2] Fig. 5). Independent peptide nucleic acid fluorescence in situ hybridization (PNA-FISH) and dual-point-paint experiments from our previous work further support *APP* gencDNAs[3] (Table 1). Critically, SMRT-seq identified 11 single-nucleotide variations that are considered pathogenic in familial AD and that were present only in our samples from individuals with SAD; none of them exist as plasmids in our laboratory.

Kim et al. compared *APP* gencDNA copy number estimates from pull-down sequencing and DISH. However, a direct comparison is not possible since the two methodologies are fundamentally different. For example, pull-downs use solution hybridization on isolated DNA, whereas DISH uses solid-phase hybridization on fixed and sorted single nuclei. Moreover, the sequences targeted are not the same. Pull-down probes target wild-type sequences for endogenous and gencDNA loci, resulting in pull-down competition. By contrast, DISH probes target only gencDNA sequences to provide greater sensitivity. Competition by

Table 1 | Summary of targeted and non-targeted *APP* PCR methods and lines of evidence that support *APP* gencDNAs and IEJs

	Method	Targeted *APP* PCR	Support for the existence of IEJs and gencDNAs	Reference
Approaches without targeted *APP* PCR				
1	RISH on IEJ 3/16	None	IEJ 3/16 RNA signal is present in human SAD brain tissue	Lee et al.[2]
2	Whole-transcriptome SMRT-seq	None	An independent commercial source identified IEJs in *APP* and other genes	Public dataset[a], Lee et al.[2] this Reply
3	Targeted RNA SMRT-seq	None	RNA pull-down that identified *APP* IEJs	Public dataset[a], Lee et al.[2]
4	DISH of gencDNAs	None	IEJ 3/16 and exon–exon junction 16/17 showed increases in neurons compared to non-neurons from the same brain from an individual with SAD and to non-diseased neurons; J20 mice containing the *APP* transgene under a PDGF-β-promoter showed increased number and size of signal compared to non-neurons and wild-type mice	Lee et al.[2]
5	Dual point-paint FISH	None	Identified *APP* CNVs of variable puncta size that were not always associated with Chr21	Bushman et al.[3]
6	PNA-FISH	None	*APP* exon copy number increases show variable signal size and shape with semiquantitative exonic probes	Bushman et al.[3]
7	Agilent SureSelect targeted pull-down	None	Identified *APP* gencDNAs in brains from individuals with SAD; contains plasmid sequence contamination	Lee et al.[2], this Reply
New #7	Agilent all-exon pull-down	None	All-exon pull-downs, with no plasmid contamination by both Vecscreen and Vecuum, contain *APP* gencDNA sequences and evidence of gencDNA UTRs and novel insertion sites	Park et al.[4], this Reply
Approaches with targeted *APP* PCR				
8	RT–PCR and Sanger sequencing	Oligo-dT primed and targeted *APP* primers	Novel *APP* RNA variants with IEJs; predominantly in neurons from individuals with SAD	Lee et al.[2]
9	Genomic DNA PCR and Sanger sequencing	Yes	Identified *APP* gencDNAs with IEJs; predominantly in neurons from individuals with SAD	Lee et al.[2]
10	Genomic DNA PCR and SMRT-seq	Yes	IEJ/gencDNAs were more prevalent in number and form in neurons from individuals with SAD compared to non-diseased neurons; identified 11 pathogenic SNVs that were present only in SAD samples	Lee et al.[2]
11	APP-751 overexpression in CHO cells	Yes	IEJ and gencDNA formation required DNA strand breakage and reverse transcriptase	Lee et al.[2]
12	Single-cell qPCR	Yes; individual exon	Intragenic exon 14 single-cell qPCR showed copy number increases in prefrontal cortical neurons over cerebellar neurons from the same brain of an individual with SAD	Bushman et al.[3]

CNV, copy number variation.

[a] The Alzheimer brain Iso-Seq dataset was generated by Pacific Biosciences, Menlo Park, California. Additional sequencing information and analysis is provided at https://downloads.pacbcloud.com/public/dataset/Alzheimer_IsoSeq_2016/.

wild-type loci reduces the efficiency of capture, which is underscored by 32% to 40% of nuclei that do not contain gencDNAs and would contribute only wild-type sequences (Lee et al., Fig. 5c, f). Moreover, a majority of gencDNA positive nuclei (62% to 73%) showed two or fewer signals (Lee et al., Fig. 5c, f) which reduced the relative representation of gencDNA loci. As IEJs do not contain the full exon sequence, there is inefficient hybridization and a lack of sequence capture and detection. This limitation is overcome by SMRT-seq (Table 1). Lastly, multiple other protocol variations exist, including tissue preparation, fixation, and hybridization conditions, which explain the hypothesized discrepancies.

Kim et al.'s third type of analysis yielded a negative result via interrogation of their own single-cell whole-genome sequencing (scWGS) data, which cannot disprove the existence of *APP* gencDNAs. An average of nine neurons from the brains of seven individuals with SAD were examined, raising immediate sampling issues required to detect mosaic *APP* gencDNAs. Kim et al. self-identified "uneven genome amplification"[1,12–14] that resulted in about 20% of their single-cell genomes having less than 10× depth of coverage[14] with potential amplification failure at one (~9% allelic dropout rate) or both alleles (~2.3% locus dropout rate)[12,14]. These limitations are compounded by potential amplification biases reflected by whole-genome amplification failure rates that may miss neuronal subtypes and/or disease states, which is especially relevant to single copies of *APP* gencDNAs that are as small as about 0.15 kb (but still detectable by DISH). Kim et al. state that the increased exonic read depth relative to introns reliably detects germline retrogene insertions in single cells from affected individuals (Kim et al., Fig. 3b); however, these data also demonstrate that increased exonic read depth is not observed in all cells—or even a majority in some cases—from the same individuals carrying the germline insertions of *SKA3* (AD3 and AD4) and *ZNF100* (AD2). These results demonstrate inherent technical limitations in the work by Kim et al. that prevent the accurate detection of even germline pseudogenes present in all cells, thus explaining an inability to detect the rarer mosaic gencDNAs produced by SGR. Kim et al.'s informatic analysis is also based on the unproven assumption that the structural features of gencDNA are shared with processed pseudogenes and LINE1 elements (Kim et al. Fig. 3a and Extended Data Fig. 1a), and possible differences could prevent straightforward detection under even ideal conditions as has been documented for LINE1[15]. These issues could explain Kim et al.'s negative results.

Considering these points, we believe that our data and conclusions supporting SGR and *APP* gencDNAs remain intact and warrant their continued study in the normal and diseased brain.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data from Park et al. were deposited in the National Center for Biotechnology Information Sequence Read Archive database under accession number PRJNA532465. Data from the newly reported full exome pull-down data sets will be provided for the *APP* locus upon request.

Code availability

The source codes of the customized algorithms are available on GitHub at https://github.com/christine-liu/exonjunction.

1. Kim, J. et al. APP gene copy number changes reflect exogenous contamination. *Nature* https://doi.org/10.1038/s41586-020-2522-3 (2020).
2. Lee, M. H. et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639–645 (2018).
3. Bushman, D. M. et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife* **4**, e05116 (2015).
4. Park, J. S. et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090 (2019).
5. Rohrback, S. et al. Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **115**, 10804–10809 (2018).
6. Cummings, J. L., Morstorf, T. & Zhong, K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res. Ther*. **6**, 37 (2014).
7. Kim, J. et al. Vecuum: identification and filtration of false somatic variants caused by recombinant vector contamination. *Bioinformatics* **32**, 3072–3080 (2016).
8. van Schendel, R., van Heteren, J., Welten, R. & Tijsterman, M. Genomic scars generated by polymerase theta reveal the versatile mechanism of alternative end-joining. *PLoS Genet*. **12**, e1006368 (2016).
9. Sfeir, A. & Symington, L. S. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci*. **40**, 701–714 (2015).
10. Splice variant case study: EGFRvIII detection in glioblastoma. https://acdbio.com/science/applications/research-areas/egfrviii (ACD, 2019).
11. Baker, A. M. et al. Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nat. Commun*. **8**, 1998 (2017).
12. Evrony, G. D. et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483–496 (2012).
13. Cai, X. et al. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep*. **8**, 1280–1289 (2014).
14. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
15. Rohrback, S., Siddoway, B., Liu, C. S. & Chun, J. Genomic mosaicism in the developing and adult brain. *Dev. Neurobiol*. **78**, 1026–1048 (2018).

Acknowledgements We thank L. Wolszon and D. Jones for manuscript editing. Research reported in this publication was supported by the NIA of the National Institutes of Health under award numbers R56AG067489 and P50AG005131 (J.C.) and NINDS R01NS103940 (Y.K.). This work was supported by non-federal funds from The Shaffer Family Foundation, The Bruce Ford & Anne Smith Bundy Foundation, and Sanford Burnham Prebys Medical Discovery Institute funds (J.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions M.-H.L., Y.K., W.J.R. and R.R. conducted laboratory experiments; C.S.L. and Y.Z. analysed sequencing data; and J.C. conceived and oversaw the experiments. G.E.K, C.S.L, and Y.Z. created figures. All authors wrote and edited the manuscript. This Reply was the work of current laboratory members.

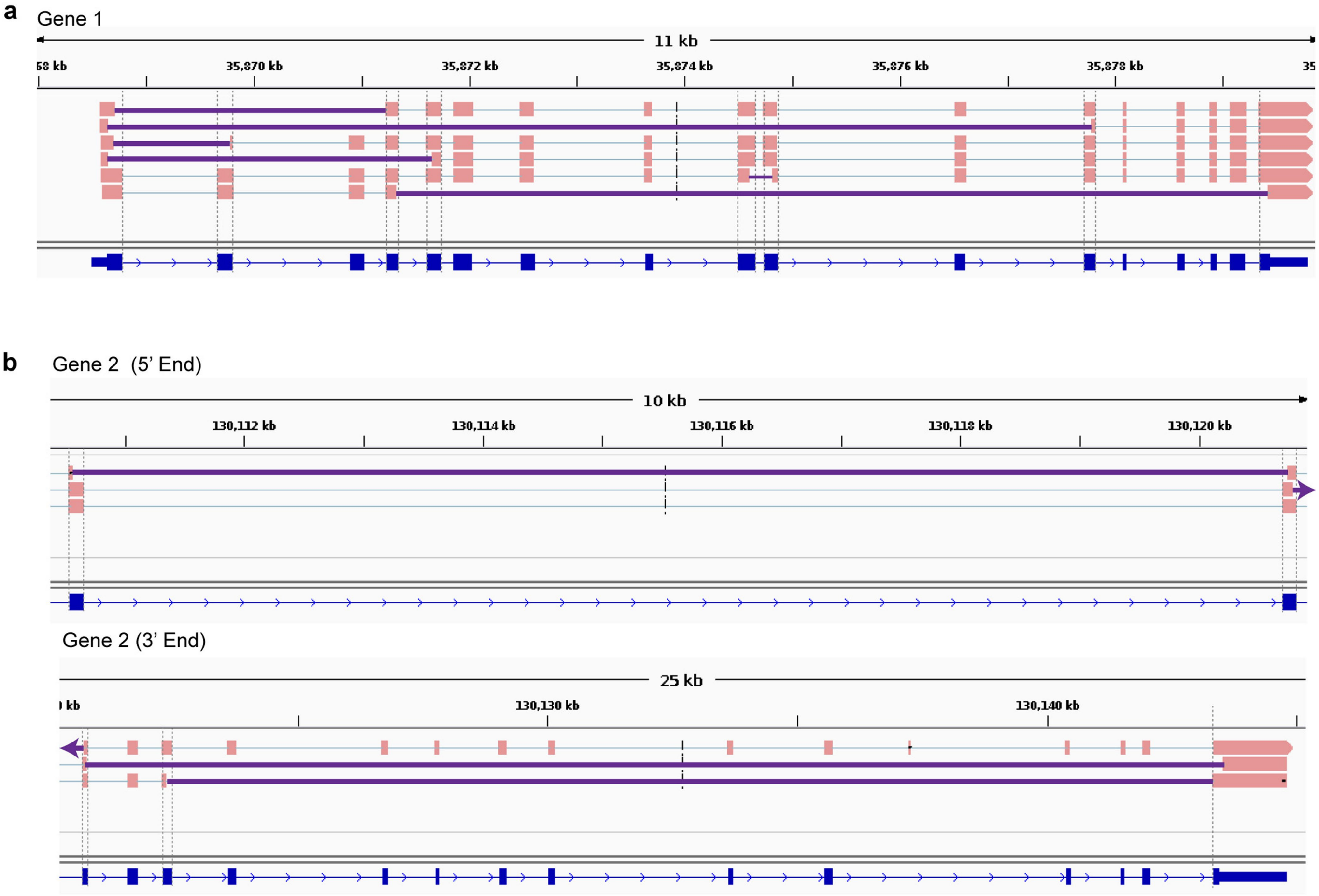
Competing interests Sanford Burnham Prebys Medical Discovery Institute has filed the following patent applications on the subject matter of this publication: (1) PCT application number PCT/US2018/030520 entitled, 'Methods of diagnosing and treating Alzheimer's disease' filed 1 May 2018, which claims priority to US provisional application 62/500,270 filed 2 May 2017; and (2) US provisional application number 62/687,428 entitled, 'Anti-retroviral therapies and reverse transcriptase inhibitors for treatment of Alzheimer's disease' filed 20 June 2018. J.C. is a co-founder of Mosaic Pharmaceuticals.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2523-2.

Correspondence and requests for materials should be addressed to J.C.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | IEJs identified from commercially available long-read transcriptome datasets in two genes other than *APP*. Sequences containing IEJs were identified and shown for gene 1 (**a**) and gene 2 (**b**). Gene 2 is shown in two parts. Grey dashed lines show ends of RefSeq exons; solid purple lines denote IEJs. All splice isoforms were examined. The Alzheimer brain Iso-Seq dataset was generated by Pacific Biosciences, Menlo Park, CA, and additional information about the sequencing and analysis is available at https://downloads.pacbcloud.com/public/dataset/Alzheimer_IsoSeq_2016/.

nature research

Corresponding author(s): Jerold Chun

Last updated by author(s): February, 21 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina sequencing of AD/MS datasets: Illumina NextSeq 500. Fastq files for Park et al. datasets were downloaded from SRA (accession PRJNA532465).
Data analysis	Sequences were aligned to the human reference genome (GRCh38) using STAR (version 2.5.3a) with the settings: --outSAMattributes All --outSJfilterCountTotalMin 1 1 1 1. Duplicate reads were marked and removed using Picard (version 2.1.1). Reads were then processed and visualized using a modified version of the R exonjunction package (https://github.com/christine-liu/exonjunction). Datasets were also analyzed using Vecuum (version 1.0.1) to confirm that APP plasmid was not detected in all of these datasets.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Fastq files of the Illumina short read sequences used in the analysis will be provided upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes indicated in figures and text were determined based on the availability of post-mortem human brain samples and the experience of the authors.
Data exclusions	No data was excluded from analysis.
Replication	All attempts at replication were successful.
Randomization	Samples were allocated randomly.
Blinding	No blinding procedure has been applied.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	All antibodies used are listed (clone number, dilution, supplier, catalog number) Rabbit monoclonal anti-NeuN antibody (27-4, 1:800, Millipore, MABN140) Alexa Fluor 488 donkey anti-rabbit IgG antibody (N/A, 1:500, Invitrogen, Ref# A21206)
Validation	These antibodies are all published and validated by immunofluorescence staining (anti-NeuN, anti-rabbit), immunohistochemistry (anti-NeuN), and Western blot (anti-NeuN). Additional validation and peer-reviewed papers are available on the manufacturer's websites.